DOCUMENT RESUME

ED 233 966 SO 015 023

TITLE	Procedural Handbook: 1981-82 Mathematics and Citizenship/Social Studies Assessments.
INSTITUTION	Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
REPORT NO	NAEP-13-MC-40
PUB DATE	Jul 83
GRANT	NIE-G-80-0003
NOTE	129p.; For related documents, see SO 015 022 and ED 186 330. Light print type marginally legible.
AVAILABLE FROM	National Assessment of Educational Progress, Educational Testing Service, Princeton, NJ 08641.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Citizenship Education; Educational Assessment; Educational Objectives; Elementary Secondary Education; *Knowledge Level; *Mathematics Education; Measures (Individuals); Program Descriptions; Questionnaires; *Social Studies; *Student Characteristics; Student Evaluation; Tables (Data)
IDENTIFIERS	*National Assessment of Educational Progress

ABSTRACT

Procedures for the National Assessment of Educational Progress (NAEP) assessments of 9-, 13-, and 17-year-old students in mathematics and citizenship/social studies are outlined. An overview of NAEP's operation is followed by descriptions of the formulation of mathematics and citizenship/social studies objectives, development and field testing of exercises, preparation of assessment materials, and data collection. Scoring of instruments is discussed in terms of quality control, training scorers, and characteristics of scoring in the specific subject areas. The discussion of data analysis focuses on exercise and group-level analysis, measurement of differences in performance over time, precautions for data analysis, and estimating variability in achievement measures. The remainder of the overview briefly covers background variables; data origins; characteristics of the respondent questionnaire, respondent data, school principal's questionnaire, and instructional program questionnaire; school and sampling records; and derived data. Appendices, which comprise more than half the document, include 15 sets of citizenship/social studies and mathematics background questions and instructional program questionnaires, background information and statistical tables on the method of reducing random variability of estimated population proportions, a glossary of NAEP terminology, and problems in inference due to the type of data and complications in analysis. Primary type of information provided by report: Procedures (overview). (KC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

NAEP Report 13-MC-40

PROCEDURAL HANDBOOK: 1981-82 Mathematics and Citizenship/Social Studies Assessments

ED232966

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM SE
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view. SO

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
T. Pratt
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Education Commission of the States
National Assessment of Educational Progress

July 1983

NIE Grant - NIE-G-80-0003

SP015-023

CONTENTS

page

FOREWORD 2

ACKNOWLEDGMENTS 4

THE 1981-82 ASSESSMENT OF MATHEMATICS AND CITIZENSHIP/SOCIAL STUDIES 5

Overview 5
Objectives 8
Mathematics Objectives 9
Citizenship/Social Studies Objectives 10
Development of Exercises 11
Mathematics Exercise Development 11
Citizenship/Social Studies Exercise Development 14
Field Testing of Exercises 16
Preparation of Assessment Materials 17
Sampling 18
Overview of the Sample Design 18
The 1981-82 Sample Design 19
Data Collection 22
Scoring 26
Training of Scorers 26
Quality Control 27
Mathematics Scoring 27
Citizenship/Social Studies Scoring 28
Data Analysis 28
Exercise Level Analysis 28
Group Level Analysis 29
Differences in Performance 30
Precautions for Data Analysis 30
Estimating Variability in Achievement Measures . 31
Quality Control 32
Data Collection 33
Scoring 33
Data Processing and Analysis 34
Background Variables 35
Standard NAEP Background Variables 35
Year-Specific NAEP Background Variables 36
Data Origins 37

Observation by Exercise Booklet Administrator . 37
Respondent Questionnaire 38
Respondent Data 38
School Principal's Questionnaire 38
Instructional Program Questionnaire 38
School Records 38
Sampling Records 39
Derived Data 39

PROCEDURAL HANDBOOK: 1981-82 Mathematics and Citizenship/Social Studies Assessments

FOREWORD

When the U.S. Office of Education was chartered in 1867, one charge to its commissioners was to determine the nation's progess in education. The National Assessment of Educational Progress (NAEP) was initiated a century later to address, in a systematic way, that charge.

Each year since 1969, National Assessment has gathered information about levels of education achievement across the country and reported its findings to the nation. NAEP surveys the education attainments of 9-year-olds, 13-year-olds and 17-year-olds in 10 learning areas: art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies and writing. Different learning areas are assessed every year, and all areas are periodically reassessed in order to measure possible changes in education achievement. National Assessment has interviewed and tested more than one million young Americans since 1969.

Learning-area assessments evolve from a consensus process. Each assessment is the product of several years of work by a great many educators, scholars and lay persons from all over the nation. Intially, these people design objectives for each subject area, proposing general goals they feel Americans should be achieving in the course of their education. After careful reviews, these objectives are given to exercise (item) writers, whose task it is to create measurement instruments appropriate to the objectives.

When the exercises have passed extensive reviews by subject-matter specialists, measurement experts and lay persons, they are administered to probability samples. The people who compose these samples are choosen in such a way that the results of their assessment can be generalized to an entire national popualtion. That is, on the basis of the performance of about 2,000 9-year-olds on a given exercise, we can make generalizations about the probable performance of all 9-year-olds in the nation.

After assessment data have been collected, scored and analyzed, National Assessment publishes reports to disseminate the results as widely as possible. Not all exercises are released for publication. Because NAEP will readminister some of the same exercises in the future to determine whether the performance level of Americans has increased, remained the same or decreased, it is essential that they not be released in order to preserve the integrity of the study.

ACKNOWLEDGMENTS

Many organizations and individuals have made substantial contributions to the mathematics and citizenship/social studies assessments. Not the least of those to be gratefully acknowleged are the adminstrators, teachers and students who cooperated so generously during the collection of the data.

Special acknowledgment must go to Jane Armstrong for coordination of development efforts for the mathematics assessment and to Ida Brooks for coordination of development efforts for the citizenship/social studies assessment. Also to be acknowledged are the substantial contributions in guiding item development and report planning made by the Mathematics Advisory Board and the Citizenship/Social Studies Advisory Board.

Administration of the 1981-82 assessments was conducted by the Research Triangle Institute, Raleigh, North Carolina. Scoring and processing were carried out by the Westinghouse Information Services, Iowa City, Iowa, and by National Assessment staff.

The preparation for the actual assessments was a collaborative effort of many members of the National Assessment staff. Special thanks go to the following people: Donald Searls for sampling and data analysis consultation; Barbara Deard and Francis Seese for preparation of item booklets and their accompanying tapescripts; and Valerie Thomas for her able assistance in all phases of preparation and analysis. The data processing and analysis of the assessment data were also a collaborative effort of National Assessment staff members. Thanks go to the following persons Fred Daniels, Gwen Edwards, John Kalk, Pam Liverpool and Cassandra Oliver. Thanks also go to Donald Phillips for coordinating the scoring of the mathematics assessment and the final data analysis for both assessments.

Beverly Anderson
Director

THE 1981-82 ASSESSMENT OF MATHEMATICS AND CITIZENSHIP/SOCIAL STUDIES

Overview

The National Assessment of Educational Progress (NAEP) completed its third assessment of mathematics and citizenship/social studies during the 1981-82 school year. Mathematics was assessed previously in 1972-73 and 1977-78; citizenship was first assessed in 1969-70 and social studies in 1971-72; both were assessed for a second time in 1975-76. Because there is a high degree of similarity between the goals considered important in citizenship and social studies education, NAEP combined the two into one comprehensive area --citizenship/social studies -- for the 1981-82 assessment.

Each assessment surveyed the achievement of American 9-, 13- and 17-year-old students using a deeply stratified, multistage probability sample. To measure changes in performance between assessments, some of the exercises administered in the previous assessments were readministered under virtually identical administrative conditions. Since 1969, National Assessment has conducted major assessments in art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies, and writing and has surveyed several other learning areas on a smaller scale. All of the major areas except career and occupational development have been reassessed one or more times. In some of the earlier assessments, 17-year-olds who were no longer in school and young adults ages 26-35 were also surveyed. Learning areas and ages assessed since 1969 are shown in Exhibit 1.

Although science was also assessed in 1981-82, this procedural handbook discusses only the mathematics and citizenship/social studies assessments. Though data were collected by National Assessment, the science assessment was sponsored by the National Science Foundation, and conducted through a contract with the University of Minnesota.

EXHIBIT 1

Learning Areas and Ages Assessed From 1969 to 1982

Assessment Year/Learning Areas	Ages Assessed*				
	9	13	17IS	17OS	Adult
1 1969-70					
Science	X	X	X	X	X
Writing	X	X	X	X	X
Citizenship	X	X	X	X	X
2 1970-71					
Reading	X	X	X	X	X
Literature	X	X	X	X	X
3 1971-72					
Music	X	X	X	X	X
Social Studies	X	X	X	X	X
4 1972-73					
Science(2)	X	X	X	X	X
Mathematics	X	X	X	X	X
5 1973-74					
Career and Occupational Development	X	X	X	X	X
Writing(2)	X	X	X	X	X
6 1974-75					
Reading(2)	X	X	X	X	
Art	X	X	X	X	
7 1975-76					
Citizenship(2)/ Social Studies(2)	X	X	X	X	
Mathematics**		X	X	X	
8 1976-77					
Science(3)	X	X	X		
Basic Life Skills**			X		
Health**					X
Energy**					X
Reading**(2)					X
Science**(3)					X

EXHIBIT 1

Learning.. (Continued)

Assessment Year/Learning Areas	Ages Assessed*				
	9	13	17IS	17OS	Adult
9 1977-78					
Mathematics(2)	X	X	X		
Consumer Skills**			X		
10 1978-79					
Art(2)	X	X	X		
Music(2)	X	X	X		
Writing(3)	X	X	X		
11 1979-80					
Reading (3)/	X	X	X	X	
Literature (2)	X	X	X	X	
Art (2)		X			
12 1980-81					
No data collection					
13 1981-82					
Mathematics(3)	X	X	X		
Citizenship/ Social Studies(3)	X	X	X		
Science**(4)	X	X	X		

NOTES:

* 17IS denotes 17-year-olds enrolled in public or private schools; 17OS denotes 17-year-olds who dropped out of school or graduated prior to the time of the assessment.

** Indicates small, special-interest probe assessments conducted on limited samples at specific ages.

( ) Shows second and subsequent assessments of an area.

Objectives

The primary goal of the National Assessment of Educational Progress is to report on the current educational status of young Americans and to monitor any changes in achievement over time. For each learning area to be assessed, NAEP asks consultants to review, revise and/or develop objectives that define the subject area. Since the objectives provide guidelines for exercise writers, consultants are asked to include examples of the knowledge, skills, understandings and attitudes to be assessed at each age level.

Education in America is a collaborative enterprise involving a great many people with widely differing philosophies. Providing information about education nationwide would be considerably easier if there were complete agreement about the means and ends of American education. The fact is, however, that Americans have conflicting and sometimes contradictory values regarding the goals of education and the means for achieving them. To have an assessment that is truly national in scope and takes into account the diversity of curricula, values and goals across the country, National Assessment employs a consensus process for developing objectives.

Several types of consultants help to develop National Assessment objectives. College and university specialists in a learning area insure that the objectives include important concepts that the schools should be teaching. Educators, including classroom teachers, curriculum supervisors and persons involved in teacher education, make sure that the objectives describe concepts, skills and attitudes that the schools should be teaching and those that they presently are teaching. Concerned citizens, parents and other interested lay persons must agree that the objectives are important for young people to achieve, are free from educational jargon and are not biased against or offensive to any groups. Consultants are selected to represent different regions of the country, community types and minority groups. They also represent a range of experience with students of different ages. Thus, the exercises used to report information about mathematics and citizenship/social studies achievement measure broad education objectives that represent a consensus of educators', subject-matter experts' and interested lay persons' opinions about what young Americans should know and be able to do. These objectives are not an attempt to mandate behavior and value systems; rather, they

represent goals that a diverse group of people have identified as desirable for young Americans to accomplish.

Mathematics Objectives

To measure the status of mathematics in 1981-82, objectives used in the 1977-78 assessment were reviewed by numerous mathematicians and mathematics educators and revised. The basic framework of the earlier objectives was retained for the 1981-82 assessment. Revisions reflect current content and trends in school mathematics.

The mathematics objectives are structured as a two-dimensional content-by-process matrix. For the 1981-82 assessment, the process dimension of the matrix includes five cognitive levels -- knowledge, skills, understanding, application and problem solving, and attitudes. The content dimension contains six categories -- numbers and numeration; variables and relationships; geometry (size, shape and position); measurement; statistics and probability; and technology.

Changes in the objectives from the second to third assessment included the following:

1. Expansion of the content by-cognitive-process matrix from four to five process levels and from five to six content levels. Attitudes were included as the fifth process level and "technology" (including computer literacy and calculators) was added as the sixth content area.

2. Addition of estimation objectives, including mental computation, computational estimation and problem solving estimation.

3. Increased emphasis on probability and statistics and problem solving.

New exercises were developed to provide coverage of the revised objectives, and exercises kept secure from previous assessments were used to measure changes in achievement across time. For further details about the objectives used to assess mathematics in 1981-82, see Mathematics Objectives, 1981-82 Assessment, (1981).

Citizenship/Social Studies Objectives

In preparation for development of objectives for the 1981-82 citizenship/social studies assessment, an initial document combining previously developed NAEP objectives in citizenship and social studies as well as other objectives developed in each of these areas was prepared in September, 1978. It was subjected to numerous revisions based on the reviews of several hundred people representing diverse regional, ethnic, cultural and educational interests.

The final objectives express a variety of perspectives regarding the goals of citizenship/social studies education. The content does not promote any particular educational, political or social viewpoint, but draws freely on all the major social science disciplines. The objectives incorporate a variety of learning approaches, but they do not prescribe any one method. Throughout, the objectives display a concern for political and cultural differences by stressing the importance of an interest in human interrelationships and a concern for basic human rights.

The National Assessment citizenship/social studies objectives are stated as general goals. Not every objective is appropriate for each of the three age levels. Good citizenship does not require 100% mastery of all of these goals. Yet achievement of the knowledge, skills, attitudes and behaviors that make up these goals should result in well-informed citizens and decision makers.

New exercises were developed to measure new citizenship/social studies objectives, and items from previous assessments were used to measure changes in achievement across time. For more information about the citizenship/social studies objectives, see Citizenship and Social Studies Objectives, 1981-82 Assessment, (1980).

Development of Exercises

Following development of objectives, exercises (items) were created to measure achievement of those objectives. Each exercise was designed so that its results could either be used alone, as an indicator of performance on a specific task, or used in conjunction with results from other exercises to give a more general picture of achievement levels.

Assessments cannot be developed directly from objectives. First, the overall structure of the item pool must be outlined. Items from previous assessments are examined to determine which will be used to detect changes in performance and whether they will be categorized according to the original objectives scheme or reorganized in a different classification system. To measure new objectives, decisions must be made as to whether all objectives will be measured, how many exercises will be used to measure each objective and the specific contents and formats appropriate to each objective.

Following the definition of the assessment's structure, items are developed to flesh out that structure. Items are developed, reviewed and revised in several iterations to ensure that they measure the intended objective, are measures of important concepts and are free from bias.

Mathematics Exercise Development

Items for the 1981-82 mathematics assessment were created through a network of development centers. Numerous consultants, including university mathematics educators, classroom teachers and lay citizens, were involved in the development, review and revisions of the items.

The development centers located at were colleges, universities or curriculum/research centers that employed a number of prominent mathematics educators willing to work together over several months to develop a portion of the exercises for the assessment. Such centers proved successful in the development of exercises for the second mathematics assessment and thus were used again for the third survey.

The centers provided numerous advantages both to NAEP and the writers. Using centers rather than conferences meant lower costs because travel was reduced. Consultants were

able to integrate work on NAEP exercise development with their other professional activities, causing a minimum of disruption and increasing their efficiency.

In the development and selection of mathematics exercises for the 1981-82 assessment, care was taken to ensure an appropriate balance of emphasis on both the content and process dimensions of the objectives matrix. This balance was achieved through attention to a questions organized according to the categories of the process dimension, questions based on the combined priorities of the interested public, mathematicians, mathematics educators and educational administrators. These questions appear on pages 16-18 of Mathematics Objectives, 1981-82 Assessment.

The content domain for the 1981-82 assessment of mathematics drew primarily from the current curriculum of elementary and secondary schools, although some projections of future mathematics emphases were acknowledged (for example, assessment of problem solving strategies and use of calculators and computers). Mathematics up to but not including calculus was included in the assessment exercises; however, a student who had no more mathematics than a first-year algebra course should have been able to answer most of the exercises.

Exercises were classified according to the six content categories, as shown in Exhibit 2. These content categories were not intended to be represented equally in the assessment. The planned weighting of exercises by content is indicated in Exhibit 2.

The five process domain categories for the third assessment are shown in Exhibit 3. Although each category suggests a type of mental process, neither objectives nor exercises fall neatly into a single process category because different students may use different processes or different combinations of processes. Thus, any system of process categories necessarily involves some arbitrary distinctions. Such a system is helpful, however, in considering the diversity of process levels possible within a given content category. The percentages in Exhibit 3 give the planned weighting of exercises by cognitive process levels.

EXHIBIT 2

Target Percentages of Exercises by Age and Content* to be Used in the 1981-82 Assessment of Mathematics

	Age 9	Age 13	Age 17
A. Numbers and numeration	40%	40%	35%
B. Variables and relationships	10	10	20
C. Shape, size and position	10	10	10
D. Measurement	15	15	10
E. Probability and statistics	5	5	10
F. Technology	10	10	10

*These percentages do not add to 100% because the attitudinal exercises are not included.

Development center consultants first conducted a logical analysis of the mathematics content area for which they were developing items. Centers used the revised objectives to make necessary changes in content or emphasis in their area. Using the secure exercises as the core of the set, new exercises were developed to either replace released exercises or to cover new topics. Centers received instructions as to the number of exercises, age overlaps and process level based on the suggested weighting of the objectives matrix.

After the first pass at exercise development was completed, the exercise set was sent out to external reviewers for comments. Their comments were reviewed by center staff, editorial changes were made and the final sets with supporting documentation were sent to NAEP for field testing.

EXHIBIT 3

Target Percentages of Exercises by Age and Process Level to be Used in the 1981-82 Assessment of Mathematics

	Age 9	Age 13	Age 17
I. Mathematical knowledge	20%	15%	15%
II. Mathematical skill	25	25	25
III. Mathematical understanding	20	25	25
IV. Mathematical application	25	30	30
V. Attitudes toward mathematics	10	5	5

In addition to reviews at the development centers and mail reviews conducted by the centers, conferences were held at NAEP to review all of the exercise sets. Eight to ten consultants, representing teachers, university specialists and the development centers, were invited to Denver to review and edit the exercises. In addition to these reviews, the exercises were mailed out to an additional group of consultants. All comments were synthesized by NAEP staff and changes were made where appropriate.

Citizenship/Social Studies Exercise Development

Early in the planning stages for the development of citizenship/social studies, it was decided to survey broadly the entire domain of citizenship and social studies as defined by the objectives. No specific parameters were placed on the number of exercises needed to measure each objective, but an attempt was made to cover all objectives. It was decided that, to the extent possible, all secure exercises from previous assessments of citizenship and social studies would be used to measure changes in achievement over time.

Several factors influenced the approach used to develop exercises for the 1981-82 assessment of citizenship/social studies. Previous developments in this subject area had been carried out through subcontracts or individual contracts with consultants. Establishing development criteria and stringently monitoring the application of the criteria had proven difficult at best during previous developments. Given the number of new objectives requiring development, the controversial nature of the subject area and the general lack of specificity of the objectives, the problems of long-distance monitoring by NAEP staff seemed even greater.

Successful exercise development in a short time called for a development procedure that could start up almost immediately, adapt as events unfolded and be closely monitored by the NAEP staff. Since staff members had been involved in all phases of the objective development and review phases, they were in the best position to assist the exercise writers and to help them translate the objectives into meaningful exercises. Based on these factors, it was decided to develop exercises using primarily local consultants who worked directly with the NAEP staff.

A draft of the objectives that could be used for exercise development was first available in January, 1979. This draft was used to develop exercises during the first of three development phases. About 900 to 1300 new exercises were needed to adequately cover the objectives, More than 800 exercises were developed by exercise writers during the first phase, from January through March 1979. During the period from August to September 1979, approximately 200 additional exercises were developed. After field testing and consultant reviews of exercises, a second development phase was initiated in August, 1979. The phase was intended to fill gaps in the exercise coverage of the objectives. Based on the losses during reviews and consultant suggestions, exercises were needed that reflected a greater diversity of experience in the United States from a multi-cultural, multi-ethnic and global perspective.

Based on decisions made by the Citizenship/Social Studies Advisory Committee and the results of review conferences held in December 1979, a third development phase occurred in the first three months of 1980, to fill more gaps in coverage.

Field Testing of Exercises

After exercises for mathematics and citizenship/social studies had been developed, critiqued and revised, they were reviewed by NAEP staff. The results of these reviews were compiled and once again the exercises were revised. Exercises that survived the reviews were field-tested in schools across the country to discover problems in wording, directions or administrative procedures and to collect exercise statistics, timing information and scoring information. Schools were selected to represent high- and low-income communities as well as more typical communities. The field tests were administered to students in at least four classrooms (approximately 100 students) at each of the ages assessed.

So that field tests closely simulated actual assessment field procedures, the students recorded their answers in the exercise booklets, directions and questions were read to students from an audio tape and National Assessment staff members, rather than classroom teachers, administered the tests. The students' responses to the exercises, as well as the administrators' reports of any field problems, helped both staff and consultants to evaluate and revise the exercises. Revised exercises were sometimes field tested again.

After exercises were field-tested, the results were reviewed by National Assessment staff and panels of subject area experts, educators and lay persons from across the country. Exercises were reviewed for age appropriateness by teachers who taught students at that age. Lay citizens, representing a variety of occupations and interests, also reviewed the exercises to check for sex or racial/ethnic bias and to consider the general importance of each exercise.

A panel of subject area specialists then worked with NAEP staff to make final selection of exercises to be used in the 1981-82 assessments of mathematics and citizenship/social studies from the pool of exercises judged acceptable by the review panels.

Preparation of Assessment Materials

A total of 35 booklets were used to assess mathematics and citizenship/social studies in 1981-82 seven for 9-year-olds and fourteen for each 13- and 17-year-olds. Six of the seven booklets for 9-year-olds contained exercises from both subject areas, and one booklet contained mathematics only. At ages 13 and 17, separate booklets were used for each subject area. For the two older ages, six booklets contained citizenship/social studies exercises, and eight contained mathematics Booklets were constructed separately for each age level, since students at different ages received different sets of exercises. Thus, exercises for 9-year-olds were not sequenced in the same order as those for 13-year-olds, and so forth.

The following constraints were observed in preparing exercise booklets:

- Each booklet contained exercises of varying difficulty so that respondents would not become bored by many easy exercises or discouraged by many difficult ones.

- Exercises could not cue other exercises. In other words, the answer to one exercise could not be contained in another exercise in the same booklet.

- Each booklet was timed so that it would take no more of a respondents' time than 45 minutes, the length of a typical class period. Booklets contained approximately 30-35 minutes of exercise time and an additional 10-15 minutes of introductory material, instructions and background questions.

- Booklets were designed to be similar, insofar as possible, with respect to the number of different objectives and subobjectives measured. Exercises measuring a particular objective were scattered throughout the booklets so that many different respondents would respond to exercises related to a particular objective.

National Assessment makes every effort to minimize difficulties connected with the testing situation so that results will be, as nearly as possible, an accurate reflection of what respondents know and can do. For example, space was provided for respondents to write their answers directly in the assessment booklets, instead of on separate answer sheets. It was felt, that this procedure would reduce

possibilities for error, especially for younger respondents. Considerable effort was also devoted to developing clear instructions for procedures the respondents should use in responding to an exercise. To minimize guessing, respondents were encouraged to select the "I don't know" response option included with most multiple-choice exercises or to write "I don't know" on the answer line for open-ended exercises if they felt they did not know the answer.

Sampling

Overview of the Sample Design

National Assessment uses a deeply stratified, three-stage national probability sample design with oversampling of low-income and rural areas. The sample is designed and selected by the staff of the Research Triangle Institute (RTI), Raleigh, North Carolina, and monitored by NAEP staff.

In the first stage of sampling, the United States is divided into geographical units of counties or groups of contiguous counties meeting a minimum population size requirement. These units, called primary sampling units (PSUs), are stratified by both region and size of community. From the list of PSUs, a sample of PSUs that represents all regions and sizes of communities is drawn (without replacement) with probability proportional to population size measures. Oversampling of low-income and rural areas is first performed at this stage by adjusting the estimated population size measures of those areas to increase sampling rates. Census Employment Survey data are used within large urban PSUs to further delineate and oversample low-income areas. Counties with high proportions of rural families are also oversampled.

In the second stage, all public and private schools within each PSU selected in the first stage are listed. Schools within each PSU are selected (without replacement) with probabilities proportional to the estimated number of age-eligibles in the school.

The third stage of sampling occurs during the data collection period. A list of all age-eligible students within each selected school is made. A simple random selection of eligible students (without replacement) is obtained, and exercise booklets are administered to the selected

students. Specially trained assessment employees select the sample and administer the booklets.

Each respondent in the sample does not have the same probability of selection because some subpopulations are over-sampled and because adjustments are made to compensate for student nonresponse and for some schools' refusal to participate. The selection probability for each individual is computed, and its reciprocal is used to weight each response in statistical calculations computed by NAEP to compensate for unequal rates of sampling and to insure proper representation in the population structure.

The number of PSUs, schools within PSUs and students within schools is determined by optimum sampling principles. That is, a sample design is selected that will achieve the maximum precision for a given level of resources.

At ages 9 and 13, almost all of the noninstitutionalized population is enrolled in public and private schools. By age 17, however, up to 15% of the population has either dropped out or graduated from secondary school. When possible, out-of-school 17-year-olds are assessed in order to provide comparable coverage of the 9-, 13- and 17-year-old populations. Assessment of out-of-school 17-year-olds and young adults ages 26-35 is quite expensive, however, and in 1981-82, funds to assess these groups were not available.

The 1981-82 Sample Design

In March 1979, primary samples were selected to serve several assessments, beginning with the 1979-80 assessment. In the first stage of sampling, counties and county-equivalent independent cities or clusters of cities were stratified by region and by size of community. A total of 1,069 primary sampling units were included in the sampling frame.

In the first stage of sampling, the PSUs were stratified by the four geographic regions defined by the Office of Business Economics, U.S. Department of Commerce (see REGION in Appendix 1). These regions are labelled Northeast, Southeast, Central and West.

Within each region, PSUs were classified into five size-of-community (SOC) categories. The sampling size of commu-

nity (SOC) variable has a year-specific definition. The size-of-community categories for the 1981-82 assessment were defined as follows:

SOC 1: PSUs corresponding to U.S. Bureau of Census Standard Metropolitan Statistical Area (SMSA) counties containing all or part of a central city of 200,000 or more population ("big city") in 1970.

SOC 2: Remaining counties in "big city" SMSAs.

SOC 3: Other counties containing all or part of a place with 25,000 or more population in 1970.

SOC 4: Counties not qualifying for SOC 1, 2 or 3 and not classified as "extreme rural" (SOC 5).

SOC 5: Counties not classified as SOC 1, 2 or 3, not having 10,000 or more total 1970 urban population, having some farm employment, and having relatively high values on an "extreme rural" index that was computed based on county labor force occupational classifications.

Thus, the design defines 20 primary strata (four geographic regions crossed with five SOC categories).

Next, a size measure was associated with each PSU within each stratum. This size measure was based on student enrollment information. PSUs that included census-defined poverty areas and census-defined extreme rural areas were identified, and the size associated with these specially defined areas was doubled to ensure adequate representation of these groups.

Before the sample was drawn, the PSUs within each stratum were ordered in a serpentine fashion by state. Increasing and then decreasing values of percent racial minorities were alternated by state.

From the ordered sampling frame, five equal size samples were selected (using a probability minimum replacement algorithm) with probability proportional to size. Each sample included 64 first-stage sampling units. Four of the samples were randomly assigned to various assessments. The fifth sample was reserved as a source of replacements for primary units that refused to participate and a possible supplementary sample for a special study.

Within the primary strata, public and private schools were listed and further stratified by the estimated number of students in a school who were eligible at each age. Small schools were clustered until they formed a large enough group to respond to the same number of exercise booklets as the larger schools in a stratum. Schools or school clusters were selected (without replacement) with probability proportional to the number of age-eligibles in the school or cluster of schools. Once schools were identified, districts were contacted to check for changes in grade range and for the existence of new schools. This information was used to revise probabilities of schools' selection. The number of PSUs and the total number of schools in which assessment sessions were conducted in 1981-82 by age are shown in Exhibit 4. Assessment sessions in 1981-82 included mathematics and citizenship/social studies booklets as well as science booklets.

EXHIBIT 4

Number of Primary Sampling Units and Total Number of Schools Participating in 1981-82 by Age

Age	No. of PSUs	No. of Schools
9	64	413
13	64	540
17	64	396

Students were then selected (without replacement) with equal probability in each sampled school. The target populations for the assessment included 9-, 13- and 17-year-olds enrolled in either public or private schools at the time of the assessment who were not functionally handicapped to the extent that they could not participate in an assessment. Definitions of the age groups are: 9-year-olds were born during calendar year 1972; 13-year-olds were born during calendar year 1968; and 17-year-olds were born from October 1, 1964 through September 30, 1965. Specific groups

excluded were non-English speaking, persons physically or mentally unable to respond, and persons in institutions or attending schools established for the physically or mentally handicapped. The number of students selected was proportional to the number of age eligibles with oversampling in low-income and rural areas.

Data Collection

A professional data collection staff from the Research Triangle Institute (RTI), Raleigh, North Carolina administered the assessment booklets. This staff was used instead of school personnel to minimize the burden on participating schools and to ensure, as far as possible, uniform administration conditions across the country. NAEP staff worked closely with the subcontractor to ensure adherence to rigorous administrative standards.

Participation in the National Assessment is voluntary; however, National Assessment makes every effort to encourage schools selected in the sample to participate in the assessment. If less than 75% of the selected 17-year-olds were available for a regular assessment session, a follow-up session was held in an effort to have a higher rate of completion by 17-year-old students. High rates of school and student cooperation were obtained in 1981-82, as shown in Exhibit 5.

EXHIBIT 5

School Cooperation Rates and Average Student Completion Rates in 1981-82 by Age

Age	Percent of Eligible Schools Participating	Average Student Completion Rates in Participating Schools
9	88.3%	90.5%
13	89.2%	85.5%
17	86.5%	74.2%

Approximately 14,000 9-year-olds, 23,000 13-year-olds and 23,000 17-year-olds participated in the 1981-82 assessment of mathematics and citizenship/social studies. Because National Assessment reports results for groups of students instead of individuals, it is not necessary for each student to answer every exercise (item). Each student completed only one exercise booklet of about 45 minutes in length. Approximately 2,000 students at each age responded to each mathematics booklet; approximately 2,000 9-year-olds and 1,200 13- and 17-year-olds responded to each citizenship/social studies booklet. The number of booklets per age and the average number of respondents assessed per booklet is indicated in Exhibit 6.

EXHIBIT 6

Number of Mathematics and Citizenship/Social Studies Booklets and Average Number of Respondents Per Booklet in 1981-82 by Age

	Age	Number of Booklets	Average Number of Respondents Per Booklet
Citizenship/ Social Studies	9	6*	1990
Mathematics	9	7*	2006
Citizenship/ Social Studies	13	6	1211
Mathematics	13	8	1970
Citizenship/ Social Studies	17	6	1125
Mathematics	17	8	2040

* Seven booklets were administered to 9-year-olds. In 6 other booklets, approximately 1/3 of the exercises surveyed citizenship/social studies and 2/3 were mathematics items. One booklet contained only mathematics exercises.

In 1981-82, all booklets were administered to groups of 10-25 students. The groups varied in size depending on the number of eligible students and on an estimate of the nonresponse rate for a particular school. Allowing variable group sizes within schools enabled National Assessment to obtain desired sample sizes in schools having characteristically low response rates. This feature also permitted last-minute modifications and adjustments to selection probabilities necessitated by enrollment changes.

Respondents within each age group were assessed at approximately the same time in the school year: 13-year-olds from October to December, 9-year-olds in January and February and 17-year-olds in March and April.

National Assessment takes steps to guarantee the anonymity of each respondent. Students' names were listed with their booklet identification number so that scoring and processing personnel could go back to the school lists for data verification -- for instance, of background information -- if necessary. These lists did not leave the schools and were destroyed six months following the assessment in a school.

Paced audio tapes were used to insure uniform assessment conditions across the country and to move respondents through the booklets at the same speed. All directions for answering exercises were read on the tape. For most of the exercises the written portions of the exercise and the response options were read aloud to reduce the effect of reading difficulties.

School officials were asked to respond to a "School Principal's Questionnaire," which included questions about the size and type of community served by the schools. In addition, school officials in schools in which citizenship/social studies exercises were administered were asked to respond to an "Instructional Program Questionnaire," which asked about citizenship/social studies programs in the school. Schools in which only mathematics booklets were administered did not receive an "Instructional Program Questionnaire". Students also provided family or personal background information through questions included in the exercise booklets. Copies of forms used to collect background information from students and school officials in the 1981-82 mathematics and citizenship/ social studies assessments are included in Appendix 1.

The assessment administrator coded for each respondent on birth date, sex, grade and racial/ethnic classification on his or her booklet. Administrators made a visual racial/ethnic identification at the time each booklet was turned in. For the 1981-82 assessments, six different racial classifications were used: American Indian or Alaskan Native, Asian or Pacific Islander, black, Hispanic heritage, white, and other or unclassified. If an administrator was unsure of a respondent's racial/ethnic group, he or she referred to the respondent's name or listened to the respondent's speech pattern to make the identification. Respondents were not verbally asked to give a racial identification for themselves by the assessment administrator; however, respondents were asked to provide this information in one of the background questions included in the assessment booklet at ages 13 and 17.

Following data collection, assessment administrators sent completed materials to the scoring contractor, Westinghouse Information Services (formerly Measurement Research Center), Iowa City, Iowa. Booklets were quality-checked to verify that correct administrative procedures were being followed by the field staff and that all materials were accounted for. Coded identification information was also checked for accuracy. Inconsistencies that could not be reconciled were sent back to the assessment administrator to be checked against the list of student names and identification numbers retained by the school for six months following the assessment.

For detailed information about the 1981-82 National Assessment data collection procedures, see Final Report-Year 13 In-School Field Operations and Data Collection Activities, National Assessment of Educational Progress. Research Triangle Park, N.C.: Research Triangle Institute, Nov., 1982.

Scoring

Scoring and computer recording of data were contracted to Westinghouse Information Services (WIS), Iowa City, Iowa, for the 1981-82 mathematics and citizenship/social studies assessments. National Assessment has found it most efficient to have scoring done by an outside contractor. By having the same contractor do both the machine scoring and the open-ended or hand-scoring, booklets to be scored did not have to be shipped to another location when different scoring methods were needed. In addition, the scoring contractor has a trained staff of scoring personnel that can be called upon and augmented when National Assessment conducts a major scoring effort.

Responses to all multiple-choice exercises were read directly by optical scanning machines. The scoring contractor employed a special staff to hand score open-ended exercises. Special training sessions were conducted by NAEP staff and WIS scoring administrators to insure that there was a high degree of uniformity in scoring procedures. Scorers were responsible for categorizing open-ended responses. They used scoring guides that defined discriptive categories of acceptable and unacceptable responses. They then coded this information into ovals that could be read by the optical scanning machine.

Training of Scorers

Scoring guides for open-ended exercises were developed using field test data. Scoring categories included likely errors and thus were useful in identifying frequently made mistakes.

National Assessment staff worked with WIS staff to train scorers. Scorers were trained to use the scoring guides by scoring samples of responses taken from arriving assessment data. Scorers initially worked as a group and discussed the appropriate categorization of each example response. They then worked individually on another set of responses. Discrepancies were resolved and explained. Once the group felt comfortable using the guides, they started categorizing the actual data. Supervisory personnel checked all work done for the first few days of a scoring effort to be sure that scoring was consistent.

Scoring for each age group of mathematics exercises began during the administration of the assessment for that age group, and it took from six to twelve weeks to complete each age. Two teams of eight scorers each scored mathematics items. One team of ten scorers categorized the citizenship/social studies responses. Open-ended citizenship/social studies exercises were given only for ages 13 and 17.

Quality Control

To further insure the quality and consistency of open-ended scoring quality-control checks were conducted. At regular intervals, randomly selected responses were drawn from the total pool of responses for an exercise and read by randomly selected scorers. Both the responses and the scorers were selected without replacement, and approximately 10% of the responses were included in the quality-control check. Scores for the quality-control readings were recorded, and the responses selected for quality control were then put back into the total pool of responses to be scored during the regular scoring the following week. Following both scorings (quality control and regular), the two scores for quality-control responses were compared. When it seemed necessary because of poor agreements of regular and quality control scorings, scorers were retrained and, on some occasions, work was rescored. The average percentage of agreement between scorers during quality control checks for mathematics was 99% at each of the three ages; for citizenship/social studies, it was 92% at age 13 and 93% at age 17.

Mathematics Scoring

For changes in performance to be measured accurately, scoring procedures and guides had to be the same for responses collected in different assessment years. The same scoring guides and procedures that had been used in 1977-78 were used again in 1981-82 to score open-ended exercises from the 1977-78 mathematics assessment that were readministered in 1981-82. When possible the same example responses used to train scorers for the 1977-78 assessment of mathematics were used to train scorers in 1981-82.

Responses from the 1977-78 assessment to open-ended mathematics exercises that were administered again in 1981-82 were randomly selected from each primary sampling unit in the country. These 1977-78 responses were rescored in 1981-82 to see if the 1981-82 scores were consistent with the scores given in 1977-78. Since agreement between scoring done in 1977-78 and that done in 1981-82 was generally 97% or more, NAEP concluded that scoring of the 1981-82 responses would be consistent with scoring done in 1977-78.

Citizenship/Social Studies Scoring

To insure the accuracy of change measures for the more complex open-ended citizenship/social studies exercises, responses from the 1975-76 assessment were rescored along with responses from the 1981-82 assessment. Scoring guides used in the 1975-76 assessment were revised and response examples from both assessment years were included to help define acceptable and unacceptable categories. These revised guides were used to score both 1975-76 and 1981-82 assessment responses.

Data Analysis

Exercise Level Analysis

The following description of NAEP data analysis procedures are presented as only one possible set of procedures appropriate for NAEP data. The basic measure of achievement in National Assessment reports is the weighted percentage of respondents who gave an acceptable response to a given exercise. This percentage is an estimate of the percentage of 9-, 13- or 17-year-olds who would have responded acceptably to a given exercise if every 9-, 13- or 17-year-old in the country were assessed. Since each exercise or exercise part is designed to measure some aspect of an objective or subobjective, the percentage of correct or acceptable responses indicates an estimated level of performance for the particular task at one point in time. A comparison of the percentage of acceptable responses on the same exercise for more than one point in time indicates whether proportions of the population able to perform the particular task are changing.

Procedures for estimating percentages of acceptable responses to mathematics and citizenship/social studies exercises are dependent on the sample design. Each individual's response is weighted and multiplied by an adjustment factor for nonresponse to reflect his or her selection probability. An estimate of the percentage of a particular age group who would have responded to an exercise acceptably if the entire age group were assessed is defined as the weighted number of acceptable responses divided by the weighted number of all the responses. A similar ratio of weights is used to estimate percentages of acceptable responses for reporting groups or subpopulations of interest.

When performing trend analysis across assessments, National Assessment has explored a number of weight smoothing procedures that are intended to reduce sampling variability in weights both between and within assessments. A summary of several procedures and details on the approach currently recommended are contained in Appendix 2.

Group Level Analysis

In addition to providing the percentage of correct or acceptable responses on each separate mathematics or citizenship/social studies exercise, National Assessment also reports the average performance across groups of similar exercises, such as the learning area as a whole, or a particular objective or subobjective. For most assessments, the metric used for summarizing results is the mean of the estimates of performance on the group of exercises, and it is called the mean percentage correct or mean percentage of acceptable responses. In the early years of NAEP the median was used; however, an experimental study of change measures sponsored by NAEP indicated that the sampling stability of the mean change was as good as or better than that of the several other measures of central tendency examined.

National Assessment reports on the achievement of various subpopulations of interest as well as for the nation as a whole. Groups are defined by region of the country, sex, race, size and type of community, level of parents' education and grade in school as well as by other characteristics. The difference between percentages of acceptable responses or averages for a reporting group and that of the entire age group (nation) on an exercise or group of exer-

cises can be used to describe the performance of any reporting group relative to the entire age group. See Appendix 3 for definitions of the standard NAEP background variables.

Differences in Performance

One of the most useful National Assessment achievement measure is differences in performance across time. Maintaining the same exercise or sets of exercises in making these comparisons provides a reasonable indicator of whether more or fewer people know or can do something judged important. To present a picture of changes in achievement for mathematics and citizenship/social studies, National Assessment describes the gains or losses on an exercise or group of exercises in terms of the differences in percentages or mean percentages of correct or acceptable responses. These differences may be used to see changes in achievement for the nation as a whole and for specific subpopulations.

Precautions for Data Analysis

- Unless the exercises summarized in the mean percentages of acceptable responses are identical, the means of one age group should not be compared to the means of another age group.

- When only a few exercises are summarized by a mean, one should be especially cautious in interpreting results, since a small set of exercises might not adequately cover the whole range of potential behaviors included under a given objective or subobjective.

- The mean should not be construed as an average test score, rather it should be interpreted literally as the arithmetic average of the percentage of acceptable responses obtained from National Assessment samples on a specific set of exercises.

Estimating Variability in Achievement Measures

National Assessment uses a national probability sample at each age level to estimate the proportion of people who would successfully complete an exercise. The particular sample selected is one of a large number of all possible samples of the same size that could have been selected with the same sample design. Since an achievement measure computed from each of the possible samples would differ from one sample to another, the standard error statistic is used as a measure of the sampling variability among achievement measures from all possible samples.

In addition to sampling variability, the standard error provides an estimate of other random error associated with the assessment. The standard error includes all nonsystematic error associated with administering specific exercises to specific students in specific situations. It also includes random differences among scorers for open-ended exercises.

In the interest of sampling and cost efficiencies, National Assessment uses a complex, stratified, multistage probability sample design. Typically, complex designs do not provide for unbiased or simple computation of sampling errors. A reasonably good approximation of standard error estimates of acceptable response percentages is obtained by applying the jackknife procedure to first-stage sampling units, or replicates, within strata. Standard errors for achievement measures such as group differences, mean percentages or mean group differences for a particular assessment year are estimated directly, and they take advantage of features of the jackknife procedure that are generic to all of these statistics. Appendices 4 and 5 contain more details about National Assessment's approach to estimating standard errors.

Quality Control

The quality control for National Assessment is largely determined by the assessment design. The assessment is intended to measure changes in performance over time and to provide products and services for users. Exercises must be sufficiently well-documented to allow reuse by NAEP or external users. Thus, printing, sampling, data collection, scoring, analysis and data processing procedures must be replicable at some time in the future. Also, the assessment must be able to detect average changes at the national level of one to two percent in the percentage of correct responses. This precision requirement places severe constraints on the amount of either sampling or nonsampling error that can be tolerated at any stage of the project. All of these factors require a much higher level of documentation and quality control than is required in a small, one-time project. While quality control procedures exist throughout the assessment, the remainder of this section focuses on quality control in data collection, scoring, data processing and analysis.

Data Collection

All data collection is performed by a specially trained field staff. The quality of data collection is monitored through:

- Field visits by NAEP and contractor supervisory staff.
- Interviews of school staff and respondents.
- Analysis of problems reported by the scoring contractor.
- Analysis of quality control questionnaires received from participating schools and districts.

Scoring

Scoring activities are divided into three major areas: receipt of data from field staff; hand scoring; and machine scoring and processing.

During receipt of data from the field staff, all materials are checked for:

- Timely receipt of all materials distributed to field staff.
- Accuracy of demographic data.
- Completeness.
- Adherence to specified procedures.
- Special problems requiring follow-up with field staff, the data collection contractor or NAEP.

During hand scoring of open-ended exercises, quality consistency is insured through the following procedures:

- Training scorers.
- Monitoring performance of scorers.
- Determining the consistency of scoring.

- Regular quality control readings of sample responses.

In machine scoring and processing, the contractor maintains a scanner error rate of less than one per hundred thousand characters through engineering checks and the following procedures:

- Tests by the contractor of all programs.
- Tests by NAEP of all programs using test data.
- Manual editing of all illegal values in data fields.

Data Processing and Analysis

To provide independent checks on data accuracy, a great deal of auditing is built into data processing and analysis procedures. Major quality control steps include:

- Manually editing and machine editing all exercise and other data file documentation.
- Manually verifying respondent data files against original contractor files.
- Verifying weight distributions from NAEP files against contractor distributions and comparing with past data and census data.
- Documenting audit trails of all processing steps and diagnostic data.
- Computing preliminary and final data independently to provide checks of results.
- Editing response ranges and consistency of data fields during contractor processing, NAEP respondent file loading, preliminary analysis and summary file loading.
- Verifying control totals for all analyses.
- Checking unusual or inconsistent patterns of results.

Background Variables

National Assessment collects background information in addition to responses to assessment exercises. Typically, exercise variables are cognitive, attitudinal or experiential questions that relate to learning area objectives. Background variables are all other information about respondents, their schools or communities, data collection conditions, etc. It is not always clear whether a particular variable ought (in some sense) to be a background or exercise variable.

Background information is collected from a number of sources. Some information comes from the exercise booklets, and some is provided by school officials responding to questionnaires. Some information is collected from observations by the test administrators, from school records or from sampling records.

Some background variables are recoded or derived from the original information gathered and are called derived variables. For example, parental education is derived from father's education and mother's education.

Some background information is collected about each individual respondent, and some is collected for each participating school or for some larger level unit.

NAEP background variables may be divided into three categories: 1) standard NAEP background variables, 2) other common NAEP background variables, and 3) year-specific NAEP background variables. The following describes these three types of background variables.

Standard NAEP Background Variables

Twelve standard background variables have a consistent definition across ages and assessments. These variables are present for almost all assessment years. These variables are:

Census Division (9 divisions)
Grades in School (collapsed into most common categories)
Race/ethnicity (3 categories)
Race/ethnicity (4 categories)
Community Size

Home Environment
Modal Grade
Parents' Education
Geographic Region (4 regions)
Sex
Size and Type of Community
Type of Community

Definitions for most standard variables can be found in Appendix 3, the Glossery.

In some cases standard NAEP variables are defined directly from the background information that has been collected. In other cases these variables are derived from other variables. Across years, the variables may be based on different category systems. For example, racial/ethnic categories have evolved over time from the simple black/ white/other categories used in 1969-70 to the current American Indian or Alaskan Native/ Asian or Pacific Islander/ black/ Hispanic heritage/white/other categories. However, by recoding data, the variables are made consistent across all years. The four category (black, white, Hispanic heritage, other)racial/ethnic variable provides a consistent variable from 1972-73 to the present, while the three category (black, white, other) racial/ethnic variable provides a consistent racial/ethnic variable for all assessments except 1970-71. In that year only, Hispanic respondents were coded as whites. In all other years Hispanic respondents were coded as other.

Year-Specific NAEP Background Variables

For each assessment, background information is collected relevant to the particular content area being assessed. For the most part, these data are collected from students in the form of specific background questions. These are either included in most or all exercise booklets or asked of school personnel through an instructional program questionnaire.

School background information on the following topics was colleted in 1981-82:

- Social studies teaching staff

- Social studies materials available in the school

- Type of materials and/or teaching approaches used to teach citizenship/social studies.

- Enrichment activities used in teaching citizenship/social studies

- Instruction given to enable student to obtain skills and knowledge in

  A. Acquiring information

  B. Using information

  C. Understanding individual development

  D. Understanding the ways human beings organize

  E. Understanding the development of the United States

No school-level instructional program questionnaire about mathematics instruction was given in 1981-82. Students responded to questions about the kinds of mathematics-related experiences that they had had. Responses to these items were used to create background variables dealing with use of the metric system, use of hand-held electronic calculators and, for age 17, the type of mathematics courses taken.

Data Origins

The data used by National Assessment in their analyses come from a variety of sources. The following describes these sources of information. Many of the forms used are found in Appendix 1.

Observation by Exercise Booklet Administrator

Respondent-level sex and race identification and administration information are provided by the field administration staff as they conduct exercise booklet administrations.

Respondent Questionnaire

Respondent-level background questions are included at the end of every exercise booklet (see Appendix 1).

Respondent Data

Respondent-level exercise data comprise the major portion of exercise booklets.

School Principal's Questionnaire

A standard school-level questionnaire is used for updating sampling records, constructing the size-and-type of community variable and providing background data (see Appendix 1). When possible, missing data are estimated from secondary sources, including Census, Office for Civil Rights, state directories and Curriculum Information Center files.

Instructional Program Questionnaire

At all three ages, a special instructional program questionnaire was administered to all sample schools that received booklets of citizenship/social studies exercises (see Appendix 1). The questionnaire asked about instructional methods and materials related to citizenship/social studies instruction.

School Records

Besides respondent's birth month and year and grade in school, school-level data are obtained from school personnel or records.

Sampling Records

Sampling records are used for a variety of identifiers, demographic variables and other data provided by the contractors from a number of sources. Except for respondent identification numbers, all variables are school-level or higher (region, community size, etc.).

Derived Data

Some variables are constructed or derived from original data fields by National Assessment staff.

LIST OF EXHIBITS

Exhibit		page
1.	Learning Areas and Ages Assessed From 1969 to 1982	6
2.	Target Percentages of Exercises by Age and Content* to be Used in the 1981-82 Assessment of Mathematics	13
3.	Target Percentages of Exercises by Age and Process Level to be Used in the 1981-82 Assessment of Mathematics	14
4.	Number of Primary Sampling Units and Total Number of Schools Participating in 1981-82 by Age	21
5.	School Cooperation Rates and Average Student Completion Rates in 1981-82 by Age	22
6.	Number of Mathematics and Citizenship/Social Studies Booklets and Average Number of Respondents Per Booklet in 1981-82 by Age	23

CONTENTS

Background Variable Questions

		AGE	PAGE
1.	Citizenship/Social Studies Background Questions.	13,17	2
2.	Citizenship/Social Studies Background Questions.	13,17	3
3.	Citizenship/Social Studies Background Questions.	9,13,17	4
4.	Mathematics Background Questions	9	5
5.	Mathematics Background Questions	13,17	6
6.	Mathematics Background Questions	17	7
7.	Mathematics Background Questions	17	8
8.	Citizenship/Social Studies and Mathematics Background Questions	13,17	9
9.	Citizenship/Social Studies and Mathematics Background Questions	13,17	10-11
10.	Citizenship/Social Studies and Mathematics Background Questions	13,17	12-13
11.	Citizenship/Social Studies and Mathematics Background Questions	9	14
12.	Citizenship/Social Studies and Mathematics Background Questions	13	15
13.	Citizenship/Social Studies and Mathematics Background Questions	17	16
14.	Citizenship/Social Studies Instructional Program Questionnaire.	9	17-24
15.	Citizenship/Social Studies Instructional Program Questionnaire.	13,17	25-32

Have you ever studied how to use any of the following? Fill in only one oval for each part.

	Yes	No	I don't know.
A. Maps	⬭	⬭	⬭
B. Globes	⬭	⬭	⬭
C. Charts	⬭	⬭	⬭
D. Tables	⬭	⬭	⬭
E. Graphs	⬭	⬭	⬭

STOP DO NOT CONTINUE UNTIL TOLD TO DO SO

A. How often do you talk about controversial topics to your TEACHERS?

- ⬭ Hardly ever
- ⬭ One day a week
- ⬭ Two days a week
- ⬭ Three or four days a week
- ⬭ More than four days a week

B. How often do you talk about controversial topics to your PARENTS?

- ⬭ Hardly ever
- ⬭ One day a week
- ⬭ Two days a week
- ⬭ Three or four days a week
- ⬭ More than four days a week

C. How often do you talk about controversial topics to your FRIENDS?

- ⬭ Hardly ever
- ⬭ One day a week
- ⬭ Two days a week
- ⬭ Three or four days a week
- ⬭ More than four days a week

DO NOT CONTINUE
UNTIL TOLD TO DO SO.

D-000002AD1D-2.3

A. How often do you talk about government or politics to your TEACHERS?

- ⬭ Hardly ever
- ⬭ One day a week
- ⬭ Two days a week
- ⬭ Three or four days a week
- ⬭ More than four days a week

B. How often do you talk about government or politics to your PARENTS?

- ⬭ Hardly ever
- ⬭ One day a week
- ⬭ Two days a week
- ⬭ Three or four days a week
- ⬭ More than four days a week

C. How often do you talk about government or politics to your FRIENDS?

- ⬭ Hardly ever
- ⬭ One day a week
- ⬭ Two days a week
- ⬭ Three or four days a week
- ⬭ More than four days a week

DO NOT CONTINUE UNTIL TOLD TO DO SO.

For each of the following questions, fill in one oval in each box.

A. The metric system uses units like centimeters, liters and kilograms. Have you used the metric system of measurement?

Yes ⬭ No ⬭ I don't know. ⬭

B. How often have you used the metric system in mathematics?

Often ⬭ Sometimes ⬭ Never ⬭ I don't know. ⬭

C. How often have you used a hand calculator?

Often ⬭ Sometimes ⬭ Never ⬭ I don't know. ⬭

D. Do you or your family own a hand calculator?

Yes ⬭ No ⬭ I don't know. ⬭

E. Does your school have hand calculators that you can use in mathematics class?

Yes ⬭ No ⬭ I don't know. ⬭

STOP DO NOT CONTINUE UNTIL TOLD TO DO SO.

5-000002692D-1

For each of the following questions, fill in one oval in each box.

A. The metric system of measurement uses units like centimeters, liters, and kilograms. How often have you used the metric system?

Often ⬭ Seldom ⬭ Never ⬭ I don't know. ⬭

B. How often do you use a hand calculator?

Almost Daily ⬭ A few times a week ⬭ Less than once a week ⬭ Once a month ⬭ Never ⬭ I don't know. ⬭

C. Does your school provide hand calculators for use in mathematics class?

Yes ⬭ No ⬭ I don't know. ⬭

DO NOT CONTINUE UNTIL TOLD TO DO SO.

For each of the following questions, fill in one oval in each box.

A. The metric system of measurement uses units like centimeters, liters, and kilograms. How often have you used the metric system of measurement?

Often ⬭ Seldom ⬭ Never ⬭ I don't know. ⬭

B. How often do you use a hand calculator?

Almost Daily ⬭ A few times a week ⬭ Less than once a week ⬭ Once a month ⬭ Never ⬭ I don't know. ⬭

C. Does your school provide hand calculators for use in mathematics classes?

Yes ⬭ No ⬭ I don't know. ⬭

D. Does your school provide hand calculators for use in other classes?

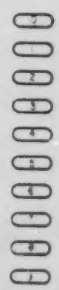
Yes ⬭ No ⬭ I don't know. ⬭

STOP DO NOT CONTINUE UNTIL TOLD TO DO SO.

5-000006ED1D-3
5-000003E919-3

Which of the following mathematics courses have you studied? Fill in one oval on each line. (If you have not studied a particular course, fill in the oval under "Not Studied".)

		Studied 1 school year	Studied ½ school year	Studied less than ½ year	Not studied	I don't know.
A.	General, Business or Consumer Mathematics	⬭	⬭	⬭	⬭	⬭
B.	Introduction to Algebra (Pre-Algebra)	⬭	⬭	⬭	⬭	⬭
C.	1st year Algebra	⬭	⬭	⬭	⬭	⬭
D.	2nd year Algebra	⬭	⬭	⬭	⬭	⬭
E.	Geometry	⬭	⬭	⬭	⬭	⬭
F.	Trigonometry	⬭	⬭	⬭	⬭	⬭
G.	Probability & Statistics	⬭	⬭	⬭	⬭	⬭
H.	Computer Programming	⬭	⬭	⬭	⬭	⬭
I.	Pre-Calculus/ Calculus	⬭	⬭	⬭	⬭	⬭



DO NOT CONTINUE UNTIL TOLD TO DO SO.

How much television did you watch yesterday?

- ⬭ None
- ⬭ 1 hour or less
- ⬭ 1 hour
- ⬭ 2 hours
- ⬭ 3 hours
- ⬭ 4 hours
- ⬭ 5 hours
- ⬭ 6 hours or more

A. Is English the language spoken most often in your home?

- ⬭ Yes
- ⬭ No

B. Is a language other than English spoken in your home?

- ⬭ Often
- ⬭ Sometimes
- ⬭ Never

28 29

0 0 0 0 0 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0

STOP DO NOT CONTINUE UNTIL TOLD TO DO SO.

28. V-000011GD1D-1.2.3
29. V-000014FD1D-1.2.3

BEST COPY AVAILABLE

Which of the following does your family have at home? (Fill in one oval on each line.)

		Have	Do not have
A.	Newspaper received regularly	⬭	⬭
B.	Magazines received regularly	⬭	⬭
C.	More than 25 books	⬭	⬭
D.	Encyclopedia	⬭	⬭
E.	Dictionary	⬭	⬭
F.	Record player	⬭	⬭
G.	A hand held calculator	⬭	⬭
H.	A computer (NOT a hand held calculator)	⬭	⬭
I.	Typewriter	⬭	⬭
J.	Vacuum cleaner	⬭	⬭
K.	Electric dishwasher	⬭	⬭

How much television did you watch yesterday?

⬭ None ⬭ 2 hours ⬭ 5 hours
⬭ 1 hour or less ⬭ 3 hours ⬭ 6 hours or more
⬭ 1 hour ⬭ 4 hours

A. How many brothers or sisters do you have who are older than you?

⬭ None ⬭ 1 ⬭ 2 ⬭ 3 ⬭ 4 ⬭ 5 ⬭ 6 or more

B. How many brothers or sisters do you have who are younger than you?

⬭ None ⬭ 1 ⬭ 2 ⬭ 3 ⬭ 4 ⬭ 5 ⬭ 6 or more

A. What is your racial background?

⬭ American Indian or Alaskan Native
⬭ Asian or Pacific Islander
⬭ Black
⬭ White
⬭ Other (Please Specify) _____

B. Is your ethnic heritage Hispanic?

⬭ No (Not Hispanic)
⬭ Yes (Mexican, Mexican-American, Chicano)
⬭ Yes, Puerto Rican
⬭ Yes, Cuban
⬭ Yes, other Spanish/Hispanic

A. Is English the language spoken most often in your home?

⬭ Yes ⬭ No

B. Is a language other than English spoken in your home?

⬭ Often ⬭ Sometimes ⬭ Never

BEST COPY AVAILABLE

How often has each of the following been used in the courses you are taking this year? (Fill in one oval on each line.)

		Never	Seldom	Fairly Often	Frequently
A.	Listening to the teacher's lecture	⬭	⬭	⬭	⬭
B.	Participating in student-centered discussions	⬭	⬭	⬭	⬭
C.	Working on a project or in a laboratory	⬭	⬭	⬭	⬭
D.	Writing essays, themes, poetry, stories	⬭	⬭	⬭	⬭
E.	Going on field trips	⬭	⬭	⬭	⬭
F.	Having individualized instruction (small groups or one-to-one with a teacher)	⬭	⬭	⬭	⬭
G.	Using teaching machines or computer-assisted instruction	⬭	⬭	⬭	⬭
H.	Watching television lectures	⬭	⬭	⬭	⬭
I.	Studying from textbooks	⬭	⬭	⬭	⬭
J.	Library or media center assignments	⬭	⬭	⬭	⬭

How much time did you spend on homework yesterday?

⬭ No homework was assigned.
⬭ I had homework but didn't do it.
⬭ Less than one hour
⬭ Between 1 and 2 hours
⬭ More than 2 hours

Which one of the following best describes your grades so far in school?

⬭ Mostly A
⬭ About half A and half B
⬭ Mostly B
⬭ About half B and half C
⬭ Mostly C
⬭ About half C and half D
⬭ Mostly D
⬭ Mostly below D

BEST COPY AVAILABLE

80. Which of the following does your family have at home? (Fill in one oval on each line.)

		Have	Do not have
A.	Newspaper received regularly	⬭	⬭
B.	Magazines received regularly	⬭	⬭
C.	More than 25 books	⬭	⬭
D.	Encyclopedia	⬭	⬭
E.	Dictionary	⬭	⬭
F.	Record player	⬭	⬭
G.	A hand held calculator	⬭	⬭
H.	A computer (NOT a hand held calculator)	⬭	⬭
I.	Typewriter	⬭	⬭
J.	Vacuum cleaner	⬭	⬭
K.	Electric dishwasher	⬭	⬭

81. How much television did you watch yesterday?

⬭ None ⬭ 2 hours ⬭ 5 hours
⬭ 1 hour or less ⬭ 3 hours ⬭ 6 hours or more
⬭ 1 hour ⬭ 4 hours

82. A. How many brothers or sisters do you have who are <u>older</u> than you?

⬭ None ⬭ 1 ⬭ 2 ⬭ 3 ⬭ 4 ⬭ 5 ⬭ 6 or more

B. How many brothers or sisters do you have who are <u>younger</u> than you?

⬭ None ⬭ 1 ⬭ 2 ⬭ 3 ⬭ 4 ⬭ 5 ⬭ 6 or more

83. A. What is your racial background?

⬭ American Indian or Alaskan Native ⬭ White
⬭ Asian or Pacific Islander ⬭ Other (Please Specify)
⬭ Black _____

B. Is your ethnic heritage Hispanic?

⬭ No (Not Hispanic)
⬭ Yes (Mexican, Mexican-American, Chicano)
⬭ Yes, Puerto Rican
⬭ Yes, Cuban
⬭ Yes, other Spanish Hispanic

84. A. Is English the language spoken most often in your home.

⬭ Yes ⬭ No

B. Is a language other than English spoken in your home?

⬭ Often ⬭ Sometimes ⬭ Never

BEST COPY AVAILABLE

How often has each of the following been used in the courses you are taking this year? (Fill in one oval on each line.)

		Never	Seldom	Fairly Often	Frequently
A.	Listening to the teacher's lecture	⬭	⬭	⬭	⬭
B.	Participating in student-centered discussions	⬭	⬭	⬭	⬭
C.	Working on a project or in a laboratory	⬭	⬭	⬭	⬭
D.	Writing essays, themes, poetry, stories	⬭	⬭	⬭	⬭
E.	Going on field trips	⬭	⬭	⬭	⬭
F.	Having individualized instruction (small groups or one-to-one with a teacher)	⬭	⬭	⬭	⬭
G.	Using teaching machines or computer-assisted instruction	⬭	⬭	⬭	⬭
H.	Watching television lectures	⬭	⬭	⬭	⬭
I.	Studying from textbooks	⬭	⬭	⬭	⬭
J.	Library or media center assignments	⬭	⬭	⬭	⬭

How much time did you spend on homework yesterday?

⬭ No homework was assigned.
⬭ I had homework but didn't do it.
⬭ Less than one hour
⬭ Between 1 and 2 hours
⬭ More than 2 hours

Which one of the following best describes your grades so far in high school?

⬭ Mostly A
⬭ About half A and half B
⬭ Mostly B
⬭ About half B and half C
⬭ Mostly C
⬭ About half C and half D
⬭ Mostly D
⬭ Mostly below D

Which one of the following best describes your present high school program?

⬭ General
⬭ Academic or college preparatory
⬭ Vocational or technical

BEST COPY AVAILABLE

Does your family get a newspaper regularly?

⬭ Yes ⬭ No ⬭ I don't know.

Does your family get any magazines regularly?

⬭ Yes ⬭ No ⬭ I don't know.

Are there more than 25 books in your home?

⬭ Yes ⬭ No ⬭ I don't know.

Is there an encyclopedia in your home?

⬭ Yes ⬭ No ⬭ I don't know.

How much school did your father complete?
(FILL IN THE ONE OVAL which best shows how much school your father completed.)

⬭ Did not complete the 8th grade

⬭ Completed the 8th grade, but did not go to high school

⬭ Went to high school, but did not graduate from high school

⬭ Graduated from high school

⬭ Some education after graduation from high school

⬭ I don't know.

Did your father graduate from a college or university?

⬭ Yes ⬭ No ⬭ I don't know.

How much school did your mother complete?
(FILL IN THE ONE OVAL which best shows how much school your mother completed.)

⬭ Did not complete the 8th grade

⬭ Completed the 8th grade, but did not go to high school

⬭ Went to high school, but did not graduate from high school

⬭ Graduated from high school

⬭ Some education after graduation from high school

⬭ I don't know.

Did your mother graduate from a college or university?

⬭ Yes ⬭ No ⬭ I don't know.

SC ⬭ ⬭

DO NOT WRITE IN THE AREA BELOW.

How much school did your father complete?
(FILL IN THE ONE OVAL which best shows how much school your father completed.)

- ⬭ Did not complete the 8th grade
- ⬭ Completed the 8th grade, but did not go to high school
- ⬭ Went to high school, but did not graduate from high school
- ⬭ Graduated from high school
- ⬭ Some education after graduation from high school

- ⬭ I don't know.

Did your father graduate from a college or university?

⬭ Yes ⬭ No ⬭ I don't know.

How much school did your mother complete?
(FILL IN THE ONE OVAL which best shows how much school your mother completed.)

- ⬭ Did not complete the 8th grade
- ⬭ Completed the 8th grade, but did not go to high school
- ⬭ Went to high school, but did not graduate from high school
- ⬭ Graduated from high school
- ⬭ Some education after graduation from high school

- ⬭ I don't know.

Did your mother graduate from a college or university?

⬭ Yes ⬭ No ⬭ I don't know.

Where did you live on your ninth birthday?

⬭ In the United States (Please specify the state or territory.)

_____

⬭ Outside the United States (Please specify the country.)

_____

⬭ I don't know.

AL ⬭
AK ⬭
AZ ⬭
AR ⬭
CA ⬭
CO ⬭
CT ⬭
DE ⬭
FL ⬭
GA ⬭
HI ⬭
ID ⬭
IL ⬭
IN ⬭
IA ⬭
KS ⬭
KY ⬭
LA ⬭
ME ⬭
MD ⬭
MA ⬭
MI ⬭
MN ⬭
MS ⬭
MO ⬭
MT ⬭
NB ⬭
NV ⬭
NH ⬭
NJ ⬭
NM ⬭
NY ⬭
NC ⬭
ND ⬭
OH ⬭
OK ⬭
OR ⬭
PA ⬭
RI ⬭
SC ⬭
SD ⬭
TN ⬭
TX ⬭
UT ⬭
VT ⬭
VA ⬭
WA ⬭
WV ⬭
WI ⬭
WY ⬭
DC ⬭
OT ⬭
BL ⬭

OC ⬭
BL ⬭

SC ⬭

DO NOT WRITE IN THE AREA BELOW

9 13

How much school did your father complete?
(FILL IN THE ONE OVAL which best shows how much school your father completed.)

- Did not complete the 8th grade
- Completed the 8th grade, but did not go to high school
- Went to high school, but did not graduate from high school
- Graduated from high school
- Some education after graduation from high school

- I don't know.

Did your father graduate from a college or university?

- Yes
- No
- I don't know.

How much school did your mother complete?
(FILL IN THE ONE OVAL which best shows how much school your mother completed.)

- Did not complete the 8th grade
- Completed the 8th grade, but did not go to high school
- Went to high school, but did not graduate from high school
- Graduated from high school
- Some education after graduation from high school

- I don't know.

Did your mother graduate from a college or university?

- Yes
- No
- I don't know.

Where did you live on your ninth birthday?

- In the United States (Please specify the state or territory.)

_____

- Outside of the United States (Please specify the country.)

_____

- I don't know.

Where did you live on your thirteenth birthday?

- In the United States (Please specify the state or territory.)

_____

- Outside of the United States (Please specify the country.)

_____

- I don't know.

AL AK AZ AR CA CO CT DE FL GA HI ID IL IN IA KS KY LA ME MD MA MI MN MS MO MT NB NV NH NJ NM NY NC ND OH OK OR PA RI SC SD TN TX UT VT VA WA WV WI WY DC OT BL

OC BL

SC

BEST COPY AVAILABLE

Year 13
Age Class 1

O M B No 1850-0083
Expiration Date 09 82

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Instructional Program Questionnaire

Citizenship/Social Studies

The purpose of this questionnaire is to provide additional information which will be used in the analysis of the citizenship/social studies data. This questionnaire should be completed by the person(s) MOST FAMILIAR with the citizenship/social studies program in your school. If you have questions about any of the following items, please contact the National Assessment District Supervisor. Thank you for your cooperation.

Please answer the following questions regarding the citizenship/social studies program in your school. The term citizenship/social studies is meant to include any instruction or activity that relates to the study of human beings and their relationships with the social and physical environments as well as the study of the civic behavior of individuals and groups. Social studies would include any courses that involve concepts from the social science disciplines of anthropology, economics, geography, history, philosophy, political science, psychology and sociology.

PLEASE USE A SOFT LEAD PENCIL.

1.	Is there a person available for the supervision and coordination of the social studies program in your school?	Yes ⬭	No ⬭
2.	Does your school provide release time for the social studies teachers to attend meetings or workshops for social studies professionals?	Yes ⬭	No ⬭
3.	Are the social studies teachers in your school involved in the textbook selection process?	Yes ⬭	No ⬭

4. Does the copyright date on the average social studies or social science textbook used in your school bear a date earlier than 1978?

Yes ⬭ No ⬭

5. Within the past three years, have funds been available in your school for supplementary social studies or social science materials and textbooks?

Yes ⬭ No ⬭

6. Are any of the following materials and/or teaching approaches used in teaching citizenship/social studies in your school?

		Yes	No
A.	Audio-Visual materials (e.g., tapes, films, filmstrips, maps, charts, graphs, tables)	⬭	⬭
B.	High Interest/Low Vocabulary Social Studies materials (e.g., workbooks, graded texts, study guides)	⬭	⬭
C.	Games and kits (commercial or teacher made)	⬭	⬭
D.	Instructional television	⬭	⬭
E.	Team teaching, contract teaching, learning centers, peer tutoring, learning packets, etc.	⬭	⬭

7. Are any of the following enrichment activities used in teaching citizenship/social studies to help students in your school?

		Yes	No
A.	Field trips	⬭	⬭
B.	Resource and/or community people	⬭	⬭
C.	Special projects (e.g., creative projects, exhibits, travel logs, school/community partnerships)	⬭	⬭
D.	Dramatic activities (e.g., plays, reenactments, stage construction, costume designing, puppeteering)	⬭	⬭
E.	Mock elections	⬭	⬭
F.	Mock political conventions	⬭	⬭

8. Does your school district have a written policy which allows the discussion of controversial issues in school?

⬭ Yes ⬭ No

9. What percentage of your teachers are teaching ONLY social studies?

⬭ None

⬭ 1-24%

⬭ 25-49%

⬭ 50-74%

⬭ 75-100%

10. During the past three years, what PERCENTAGE of your ENTIRE teaching staff (including teachers in areas such as art, reading, music, mathematics, science, physical education, etc.) have participated in PLANNING inservice activities for your school or district?

- ⬭ None participated in planning
- ⬭ 1-9% participated in planning
- ⬭ 10-24% participated in planning
- ⬭ 25-49% participated in planning
- ⬭ 50-99% participated in planning
- ⬭ 100% participated in planning

11. During the past three years, what PERCENTAGE of your ENTIRE teaching staff (including teachers in areas such as art, reading, music, mathematics, science, physical education, etc.) have received inservice TRAINING in CITIZENSHIP/SOCIAL STUDIES from your school or district?

- ⬭ No citizenship/social studies inservice training offered
- ⬭ None received training
- ⬭ 1-9% received training
- ⬭ 10-24% received training
- ⬭ 25-49% received training
- ⬭ 50-99% received training
- ⬭ 100% received training

12. For each of the following areas, indicate whether or not the students in your school are given instruction in that area or are provided with activities that would help them develop knowledge and skills in that area.

A. ACQUIRING INFORMATION

		A part of instruction or activity	NOT a part of instruction or activity
1.	Use of the senses to obtain information	⬭	⬭
2.	Use of a variety of sources to obtain information	⬭	⬭
3.	Use of a variety of techniques (e.g., interviews, polls, surveys)	⬭	⬭

B. USING INFORMATION

		A part of instruction or activity	NOT a part of instruction or activity
1.	Organizing information (e.g., arranging, grouping, relating, analyzing)	⬭	⬭
2.	Applying information (e.g., inferring, predicting, testing alternatives)	⬭	⬭
3.	Decision making and problem solving	⬭	⬭
4.	Critically evaluating information	⬭	⬭

12. (Continued)

C. UNDERSTANDING INDIVIDUAL DEVELOPMENT

		A part of instruction or activity	NOT a part of instruction or activity
1.	Examining the individual's beliefs, values and behaviors	⬭	⬭
2.	Individual development (setting goals, determining consequences of goals)	⬭	⬭
3.	Graphic and oral communication	⬭	⬭
4.	Interpersonal communication	⬭	⬭
5.	Interaction in groups	⬭	⬭
6.	Relating to people of different cultural perspectives	⬭	⬭

D. UNDERSTANDING THE WAYS HUMAN BEINGS ORGANIZE

		A part of instruction or activity	NOT a part of instruction or activity
1.	Ways in which people are interrelated (e.g., economically, environmentally, historically, politically, culturally, etc.)	⬭	⬭
2.	Organizations of human societies (e.g., nature of institutions and groups)	⬭	⬭

BEST COPY AVAILABLE

12. (Continued)

D. UNDERSTANDING THE WAYS HUMAN BEINGS ORGANIZE

(Continued)

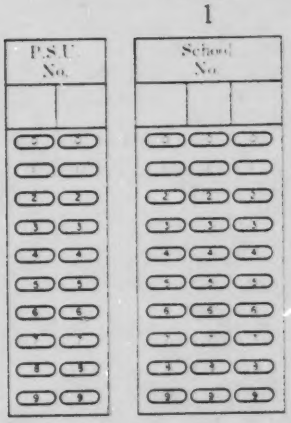
		A part of instruction or activity	NOT a part of instruction or activity
3.	Relationships between individuals and groups (e.g., roles in groups, effect of groups on individuals)	⬭	⬭
4.	Relationships between groups (e.g., cooperation and conflict among groups)	⬭	⬭
5.	Global concerns (e.g., food, population, disease, racism, conflict, human rights)	⬭	⬭
6.	Human rights worldwide (e.g., acceptance and appreciation of universal differences, equitable treatment, consideration for interest of others)	⬭	⬭

E. UNDERSTANDING THE DEVELOPMENT OF THE UNITED STATES

		A part of instruction or activity	NOT a part of instruction or activity
1.	Principles and purposes of the U.S. government	⬭	⬭
2.	Organization and operations of U.S. government (federal, state, local)	⬭	⬭

THANK YOU FOR YOUR COOPERATION.

Age Class 1



School Name:_____

Year 13
Age Classes 2 and 3

O.M.B. No. 1850-0083
Expiration Date 09/82

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Instructional Program Questionnaire

Citizenship/Social Studies

The purpose of this questionnaire is to provide additional information which will be used in the analysis of the citizenship/social studies data. This questionnaire should be completed by the person(s) MOST FAMILIAR with the citizenship/social studies program in your school. If you have questions about any of the following items, please contact the National Assessment District Supervisor. Thank you for your cooperation.

Please answer the following questions regarding the citizenship/social studies program in your school. The term citizenship/social studies is meant to include any instruction or activity that relates to the study of human beings and their relationships with the social and physical environments as well as the study of the civic behavior of individuals and groups. Social studies would include any courses that involve concepts from the social science disciplines of anthropology, economics, geography, history, philosophy, political science, psychology and sociology.

PLEASE USE A SOFT LEAD PENCIL

1.	Is there a person available for the supervision and coordination of the social studies program in your school?	Yes ⬭	No ⬭
2.	Does your school provide release time for the social studies teachers to attend meetings or workshops for social studies professionals?	Yes ⬭	No ⬭
3.	Are the social studies teachers in your school involved in the textbook selection process?	Yes ⬭	No ⬭

4. Does the copyright date on the average social studies or social science textbook used in your school bear a date earlier than 1978?

Yes ⬭ No ⬭

5. Within the past three years, have funds been available in your school for supplementary social studies or social science materials and textbooks?

Yes ⬭ No ⬭

6. Are any of the following materials and/or teaching approaches used in teaching citizenship/social studies in your school?

		Yes	No
A.	Audio-Visual materials (e.g., tapes, films, filmstrips, maps, charts, graphs, tables)	⬭	⬭
B.	High Interest/Low Vocabulary Social Studies materials (e.g., workbooks, graded texts, study guides)	⬭	⬭
C.	Games and kits (commercial or teacher made)	⬭	⬭
D.	Instructional television	⬭	⬭
E.	Team teaching, contract teaching, learning centers, peer tutoring, learning packets, etc.	⬭	⬭

7. Are any of the following enrichment activities used in teaching citizenship/ social studies to help students in your school?

		Yes	No
A.	Field trips	⬭	⬭
B.	Resource and/or community people	⬭	⬭
C.	Special projects (e.g., creative projects, exhibits, travel logs, school/community partnerships)	⬭	⬭
D.	Dramatic activities (e.g., plays, reenactments, stage construction, costume designing, puppeteering)	⬭	⬭
E.	Mock elections	⬭	⬭
F.	Mock political conventions	⬭	⬭

8. Does your school district have a written policy which allows the discussion of controversial issues in school?

⬭ Yes ⬭ No

9. What is the average number of years of teaching experience for the social studies teachers in your school?

⬭ Less than 1 year

⬭ 1-2 years

⬭ 2-3 years

⬭ 4-5 years

⬭ More than 5 years

10. What percentage of your social studies teachers are certified to teach social studies?

- ⬭ None
- ⬭ 1-24%
- ⬭ 25-49%
- ⬭ 50-74%
- ⬭ 75-100%

11. During the past three years, what PERCENTAGE of your ENTIRE teaching staff (including teachers in areas such as art, reading, music, mathematics, science, physical education, etc.) have participated in PLANNING inservice activities for your school or district?

- ⬭ None participated in planning
- ⬭ 1-9% participated in planning
- ⬭ 10-24% participated in planning
- ⬭ 25-49% participated in planning
- ⬭ 50-99% participated in planning
- ⬭ 100% participated in planning

12. During the past three years, what PERCENTAGE of your ENTIRE teaching staff (including teachers in areas such as art, reading, music, mathematics, science, physical education, etc.) have received inservice TRAINING in CITIZENSHIP/SOCIAL STUDIES from your school or district?

- ⬭ No citizenship/social studies inservice training offered
- ⬭ None received training
- ⬭ 1-9% received training
- ⬭ 10-24% received training
- ⬭ 25-49% received training
- ⬭ 50-99% received training
- ⬭ 100% received training

13. For each of the following areas, indicate whether or not the students in your school are given instruction in that area or are provided with activities that would help them develop knowledge and skills in that area.

A. ACQUIRING INFORMATION

		A part of instruction or activity	NOT a part of instruction or activity
1.	Use of the senses to obtain information	⬭	⬭
2.	Use of a variety of sources to obtain information	⬭	⬭
3.	Use of a variety of techniques (e.g., interviews, polls, surveys)	⬭	⬭

B. USING INFORMATION

		A part of instruction or activity	NOT a part of instruction or activity
1.	Organizing information (e.g., arranging, grouping, relating, analyzing)	⬭	⬭
2.	Applying information (e.g., inferring, predicting, testing alternatives)	⬭	⬭
3.	Decision making and problem solving	⬭	⬭
4.	Critically evaluating information	⬭	⬭

13. (Continued)

C. UNDERSTANDING INDIVIDUAL DEVELOPMENT

		A part of instruction or activity	NOT a part of instruction or activity
1.	Examining the individual's beliefs, values and behaviors	⬭	⬭
2.	Individual development (setting goals, determining consequences of goals)	⬭	⬭
3.	Graphic and oral communication	⬭	⬭
4.	Interpersonal communication	⬭	⬭
5.	Interaction in groups	⬭	⬭
6.	Relating to people of different cultural perspectives	⬭	⬭

D. UNDERSTANDING THE WAYS HUMAN BEINGS ORGANIZE

		A part of instruction or activity	NOT a part of instruction or activity
1.	Ways in which people are interrelated (e.g., economically, environmentally, historically, politically, culturally, etc.)	⬭	⬭
2.	Organizations of human societies (e.g., nature of institutions and groups)	⬭	⬭

13. (Continued)

D. UNDERSTANDING THE WAYS HUMAN BEINGS ORGANIZE (Continued)

		A part of instruction or activity	NOT a part of instruction or activity
3.	Relationships between individuals and groups (e.g., roles in groups, effect of groups on individuals)	⬭	⬭
4.	Relationships between groups (e.g., cooperation and conflict among groups)	⬭	⬭
5.	Global concerns (e.g., food, population, disease, racism, conflict, human rights)	⬭	⬭
6.	Human rights worldwide (e.g., acceptance and appreciation of universal differences, equitable treatment, consideration for interest of others)	⬭	⬭

E. UNDERSTANDING THE DEVELOPMENT OF THE UNITED STATES

		A part of instruction or activity	NOT a part of instruction or activity
1.	Principles and purposes of the U.S. government	⬭	⬭
2.	Organization and operations of U.S. government (federal, state, local)	⬭	⬭

13. (Continued)

E. UNDERSTANDING THE DEVELOPMENT OF THE UNITED STATES (Continued)

		A part of instruction or activity	NOT a part of instruction or activity
3.	Political decision making in the U.S.	⬭	⬭
4.	Electoral processes in the U.S.	⬭	⬭
5.	Legal system in the U.S.	⬭	⬭
6.	Rights of individuals in the U.S.	⬭	⬭
7.	Civil and criminal justice systems in the U.S.	⬭	⬭
8.	Support of justice and rights for all individuals	⬭	⬭
9.	Economics in the U.S.	⬭	⬭

14. FOR HIGH SCHOOLS ONLY:

How many carnegie units are students required to have in citizenship/ social studies curriculum for graduation in your school district?

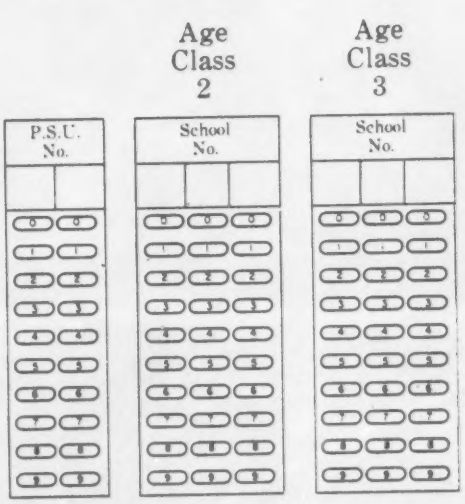
⬭ More than one carnegie unit

⬭ One carnegie unit

⬭ One-half carnegie unit

⬭ Less than one-half carnegie unit

⬭ None required



THANK YOU FOR YOUR COOPERATION.

School Name: _____

APPENDIX 2

ADJUSTMENT OF RESPONDENT WEIGHTS BY SMOOTHING TO REDUCE RANDOM VARIABILITY OF ESTIMATED POPULATION PROPORTIONS

Background

As noted elsewhere, a weight is assigned to every individual who responds to an exercise administered in an assessment. The weight is the reciprocal of the probability of selection of the individual with adjustment for nonresponse. It estimates the number of similar people that that individual represents in the age population. The sum of the weights of all individuals at an age level responding to an exercise is an estimate of the total number of people in that age population in the year that the exercise was assessed. Similarly, the sum of weights for all individuals who took the exercise and who also are members of some demographic category (such as blacks) gives an estimate of the number of people in the age population, for the year, who are also members of the category. The ratio of the two totals estimates the proportional representation of the demographic category in the age population for the given year.

For each of the assessed age populations in each assessment year, separate estimates of the proportional representation of the various demographic subgroups are provided by each booklet administered to that age group in that year. Due to random sampling variability, the estimates of population proportions for a given year based on single booklets administered in the year will vary. In addition to whatever trends in population proportions over time that might exist, there is also random sampling variation in these proportions from year to year.

It is desirable to reduce the random variability of population proportions as much as possible, since this variability has an effect on performance estimates. For example, the percentage of acceptable responses for an age group is a function of the relative proportions of high-performing and

low-performing groups. If the relative proportions of these groups are very different in different assessments due to sampling variability, then a portion of the change in percentage of acceptable responses for an age group could be attributable to yearly sampling difference in the relative proportions of high- and low-achieving groups.

In addition to reporting performance estimates for an age group as a whole, National Assessment also reports performance for various subpopulations, such as whites or blacks. Because variability of subgroups within these subpopulations (such as white males and females within the white subpopulation) influence the performance estimates for the subpopulations, it is desirable that fluctuations of proportions of all subgroups of each subpopulation be reduced as much as possible.

For each age and year, each of the various booklets administered will provide estimates of a given population proportion. Since these estimates are subject to booklet to booklet variability, a better estimate of the population proportion, which will have reduced variability, is obtained by combining the information from all booklets. However, these proportions vary from year to year due to random sampling variability or systematic differences in sampling procedures. An even better estimate of population proportions for any single year can be obtained by smoothing the proportions over several assessment years. The word "smoothing" is used here in the sense of fitting a smooth curve to a sequence of numbers by robust/resistant procedures. Smoothing estimates of population proportions reduces a large portion of the sampling variability while preserving, as far as possible, actual trends occurring in the age population.

After the population proportions have been smoothed, adjusted weights are derived for the assessed individuals so that the population proportions computed using the adjusted weights are equal to the smoothed proportions. The adjusted weights are then used for all analyses, such as estimation of performance.

The result of the smoothing and weight-adjusting process is that both adjusted performance estimates and changes in those estimates appear to be somewhat less susceptible to sampling variability, both across and within years.

APPENDIX 2

Smoothing Procedures Used by National Assessment

The most direct way to smooth proportions is first to classify people into mutually exclusive multi-way cells on the basis of their membership in categories of various important variables and then to smooth the proportions within each of the resulting multi-way cells across years. Unfortunately, this procedure tends to produce a large number of cells with few people and consequently quite unstable estimates of smoothed proportions.

To circumvent this difficulty, National Assessment has utilized various smoothing procedures since the 1976-77 assessment. Each of these procedures, which are all basically weighting-class adjustments applied independently to each age, is designed to control, to varying degrees, fluctuations in certain key subgroups while avoiding, as much as possible, instabilities due to small cells.

The procedure used for the 1976-77 assessment was a weighting class adjustment applied independently to each age and reporting variable (nation, region, sex, etc.). The details of the procedure are given in Appendix B of Technical Report 08-S-21: Three Assessments of Science, 1969-77: Technical Summary. While this procedure performs well, it is complicated and requires large amounts of time and computer resources to implement. By independently smoothing proportions within each reporting variable, it was possible to produce good estimates of the marginal proportions of people within each category of the variable, while disturbing as little as possible the relationships between other reporting variables within the adjusted variable. However, this meant that each individual had a different adjusted weight for each reporting variable under consideration. While this presents no problem for the estimation of performance within a reporting variable, the multiplicity of weights definitely complicates analyses, such as regression, that involve several variables.

APPENDIX 2

Because of the complexity of the procedure used in 1976-77, a different and simpler procedure was adopted in 1977-78. This procedure is detailed in Appendix F of Report 09-MA-40, Procedural Handbook: 1977-78 Mathematics Assessment. The 1977-78 smoothing procedure produced a single adjusted weight for each individual, and hence greatly reduced the complexity of subsequent analyses of performance data. The 1977-78 procedure involved applying a weighting class adjustment independently to each age. The weighting classes, which were different at each age, consisted of individuals who were alike on certain demographic characteristics and who would be expected to have similar educational achievement characteristics. There were around seventy adjustment cells used for each age.

Although the 1977-78 procedure produced acceptable results, National Assessment in 1978-79 adopted yet another procedure which we believe has the best characteristics of the three procedures used. The new (1978-79) procedure, which is detailed below, has several advantages.

1. It produces a single adjusted weight for each individual.

2. It affords good control on the distribution of proportions of certain key variables.

3. It tends to produce stable performance estimates.

4. It is the easiest to implement.

Even though adjusted weights using this procedure differ slightly from the corresponding adjusted weights from the other procedures, we intend to use weights obtained using this procedure for all future analyses of data assessed in earlier years. This is simply because we believe these weights to be the best available.

APPENDIX 2

The Current Smoothing Procedure

The first step in the 1978-79 smoothing procedure involved the partitioning of the population of age class eligibles into the six smoothing cells given in Exhibit 2.1 The same cells were used for all ages.

EXHIBIT 2.1

Smoothing Cells Used for the 1978-79 Procedure

Cell	Race	Region	Community Size (CS)
1	White	All	Big City + Fringe (BC + FR)
2	White	All	Medium City (MC)
3	White	All	Small Places (SP)
4	Black	SE	All
5	Black	NOT SE	All
6	Other	All	All

Then, for each age and every year, the proportion of the population in each of the cells was estimated. For a given age and year, the proportion of the population in a particular cell was computed as the sum of weights of all respondents assessed in the given year who were of the specified age and who belonged in the cell, divided by the total of the weight of all respondents of the given age assessed in that year. Proportions for 9-, 13- and 17-year-olds are shown in Exhibits 2.2A, B and C, respectively.

Each of the six cells was composed of a sequence of estimated population proportions corresponding to the various years of assessment. Each such sequence of proportions was then smoothed by fitting robust/resistant lines. Using data from the Census and Current Population Surveys, trends in enrollment by age and race and by age and region were obtained. The data from these surveys were adjusted to correspond with NAEP definitions as much as possible. The resistant lines within the smoothing cells were constrained to satisfy the trends from the Census and CPS data.

The final step in the smoothing procedure was to adjust the respondents' weights to be consistent with the smoothed proportions. Since each respondent takes only one booklet, the weight adjustments were done independently for each booklet. For a given age, year and booklet, population proportions using the original weights were obtained for each of the smoothing cells. Then the weights of all respondents of a given cell were multiplied by the ratio of the smoothed cell proportion to the proportion using the original weights. This produced the adjusted weights which are used in all analyses.

APPENDIX 2

EXHIBIT 2.2A

Unsmoothed Proportions by Smoothing Cell and Year
for Nine-Year-Olds

CELL	1	2	3	4	5	6
RACE	WHITE	WHITE	WHITE	BLACK	BLACK	OTHER
REGION	ALL	ALL	ALL	SE	NOT SE	ALL
CS	BC+FR	MC	SP	ALL	ALL	ALL
YEAR						
69-70	.33545	.11333	.36797	.053937	.066117	.063282
70-71*	.32532	.12165	.35154	.058040	.077328	.066184
71-72	.31743	.10852	.38149	.060266	.074271	.058037
72-73	.32555	.10957	.35378	.062283	.081864	.067044
73-74	.31524	.13755	.34419	.056249	.089667	.057184
74-75	.28359	.12422	.39229	.062768	.081998	.055204
75-76	.27167	.11434	.41492	.050450	.059552	.089095
76-77	.27819	.11126	.42390	.057192	.069709	.059827
77-78	.33610	.06958	.38770	.070198	.066873	.069632
78-79	.28557	.08517	.39040	.083710	.077259	.077959
79-80	.23366	.11382	.45130	.042650	.073036	.085610
81-82	.41792	.05381	.29220	.079516	.078769	.077786

NOTES:
* In 1970-71, Hispanics were included in the "White" category; they are in "Other" for all other assessments.

APPENDIX 2

EXHIBIT 2.2B

Unsmoothed Proportions by Smoothing Cell and Year
for Thirteen-Year-Olds

CELL	1	2	3	4	5	6
RACE	WHITE	WHITE	WHITE	BLACK	BLACK	OTHER
REGION	ALL	ALL	ALL	SE	NOT SE	ALL
CS	BC+FR	MC	SP	ALL	ALL	ALL
YEAR						
69-70	.31308	.13724	.35597	.059331	.073373	.061090
70-71*	.31105	.10757	.37370	.058330	.088538	.060869
71-72	.35923	.11459	.36682	.046199	.056091	.057165
72-73	.33276	.11460	.36012	.052248	.074108	.066245
73-74	.32757	.12686	.36269	.050682	.075044	.057221
74-75	.29797	.11200	.40792	.061019	.068247	.052903
75-76	.24279	.10656	.47169	.047466	.046586	.084974
76-77	.27905	.09100	.42018	.050036	.108012	.051784
77-78	.36593	.06914	.36030	.065470	.065295	.073942
78-79	.33351	.08365	.33756	.075508	.087461	.082397
79-80	.22430	.16140	.40484	.040344	.090701	.078483
81-82	.40147	.07650	.29938	.059464	.082718	.080481

NOTES:
* In 1970-71, Hispanics were included in the "White" category; they are in "Other" for all other assessments.

APPENDIX 2

EXHIBIT 2.2C

Unsmoothed Proportions by Smoothing Cell and Year
for Seventeen-Year-Olds

CELL	1	2	3	4	5	6
RACE	WHITE	WHITE	WHITE	BLACK	BLACK	OTHER
REGION	ALL	ALL	ALL	SE	NOT SE	ALL
CS	BC+FR	MC	SP	ALL	ALL	ALL
YEAR						
69-70	.33826	.15743	.37913	.036070	.061890	.027287
70-71*	.37123	.12382	.34146	.045810	.067916	.049850
71-72	.35383	.11700	.37216	.047010	.059455	.050627
72-73	.33776	.11225	.37505	.048503	.071475	.055038
73-74	.35497	.14873	.34664	.043443	.066038	.040246
74-75	.34387	.11761	.37817	.057713	.063378	.039341
75-76	.25909	.11842	.46370	.043622	.052995	.062246
76-77	.31808	.11427	.40588	.043133	.072569	.046115
77-78	.40073	.06226	.36979	.050924	.060797	.055582
78-79	.36274	.09967	.35407	.052836	.059485	.071276
79-80	.24808	.17037	.40955	.039871	.072145	.060060
81-82	.38273	.04901	.30745	.064539	.099346	.096929

NOTES:

* In 1970-71, Hispanics were included in the "White" category; they are in "Other" for all other assessments.

APPENDIX 2

Adjustment of Weights by Users

Smoothed population proportions for 1969-70 through 1979-80 are given in Exhibits 2.3A, B and C for 9-, 13- and 17-year-olds, respectively. These are the proportions used in analyses involving 1979-80 reading/literature data and other NAEP analyses conducted between late 1980 and late 1982. For example, the smoothed population proportion of 9-year-olds in smoothing cell two (whites in medium cities) for 1972-73 is .11518. Note that the 1970-71 entries are blank. In that assessment, Hispanics were included in the "white" classification rather than "other" or a separate category. Consequently, smoothed proportions have not been used by National Assessment for analyses of the 1970-71 data.

To adjust respondent weights to be consistent with the smoothed proportions, the following procedure is followed:

1. For each booklet, classify the respondents according to smoothing cell and obtain the raw population proportions for each cell. For example, the raw proportion for a booklet of 9-year-olds in smoothing cell four is the total of the weights of all 9-year-olds in the booklet who are Black and in the Southeast region, divided by the total of the weights of all respondents to the booklet.

2. For each booklet and smoothing cell, obtain a weight adjustment factor as the ratio of the smoothed population proportion (for the appropriate age, year and smoothing cell) over the raw population proportion.

3. The adjusted weight for an individual is the product of that individual's original weight and the appropriate adjustment factor.

APPENDIX 2

EXHIBIT 2.3A

Smoothed Proportions by Smoothing Cell and Year
for Nine-Year-Olds from 1969-70 to 1979-80

CELL	1	2	3	4	5	6
RACE	WHITE	WHITE	WHITE	BLACK	BLACK	OTHER
REGION	ALL	ALL	ALL	SE	NOT SE	ALL
CS	BC+FR	MC	SP	ALL	ALL	ALL
YEAR						
69-70	.32930	.12576	.35456	.053544	.074486	.06235
70-71*						
71-72	.32322	.11772	.36472	.056238	.074252	.06385
72-73	.31650	.11518	.37200	.056790	.074930	.06460
73-74	.30977	.11265	.37928	.057342	.075608	.06535
74-75	.30305	.11011	.38656	.057894	.076286	.06610
75-76	.29632	.10757	.39385	.058447	.076963	.06685
76-77	.28960	.10503	.40113	.058999	.077641	.06760
77-78	.28287	.10249	.40841	.059551	.078319	.06835
78-79	.27615	.09996	.41569	.060103	.078997	.06910
79-80	.26943	.09742	.42298	.060656	.079674	.06985

NOTES:
* 1970-71 data omitted because of non-standard racial/ethnic definitions.

APPENDIX 2

EXHIBIT 2.3B

Smoothed Proportions by Smoothing Cell and Year
for Thirteen-Year-Olds from 1969-70 to 1979-80

CELL	1	2	3	4	5	6
RACE	WHITE	WHITE	WHITE	BLACK	BLACK	OTHER
REGION	ALL	ALL	ALL	SE	NOT SE	ALL
CS	BC+FR	MC	SP	ALL	ALL	ALL
YEAR						
69-70	.32002	.13088	.37030	.051270	.067330	.06020
70-71*						
71-72	.32792	.11055	.37793	.052407	.069393	.06180
72-73	.32317	.10982	.38101	.052495	.070905	.06260
73-74	.31841	.10910	.38409	.052582	.072418	.06340
74-75	.31366	.10837	.38717	.052669	.073931	.06420
75-76	.30890	.10764	.39026	.052756	.075444	.06500
76-77	.30415	.10691	.39334	.052844	.076956	.06580
77-78	.29940	.10618	.39642	.052931	.078469	.06660
78-79	.29464	.10546	.39950	.053018	.079982	.06740
79-80	.28989	.10473	.40258	.053105	.081495	.06820

NOTES:
* 1970-71 data omitted because of non-standard racial/ethnic definitions.

APPENDIX 2

EXHIBIT 2.3C

Smoothed Proportions by Smoothing Cell and Year
for Seventeen-Year-Olds from 1969-70 to 1979-80

CELL	1	2	3	4	5	6
RACE	WHITE	WHITE	WHITE	BLACK	BLACK	OTHER
REGION	ALL	ALL	ALL	SE	NOT SE	ALL
CS	BC+FR	MC	SP	ALL	ALL	ALL
YEAR						
69-70	.34050	.14471	.36862	.04148	.058110	.04658
70-71*						
71-72	.35766	.11994	.37040	.044424	.059726	.04782
72-73	.35192	.11939	.37379	.045060	.061370	.04844
73-74	.34618	.11884	.37718	.045697	.063013	.04906
74-75	.34044	.11829	.38057	.046333	.064657	.04968
75-76	.33470	.11774	.38396	.046969	.066301	.05030
76-77	.32896	.11718	.38736	.047606	.067944	.05092
77-78	.32322	.11663	.39075	.048242	.069588	.05154
78-79	.31748	.11608	.39414	.048878	.071232	.05216
79-80	.31174	.11553	.39753	.049514	.072876	.05278

NOTES:
* 1970-71 data omitted because of non-standard racial/ethnic definitions.

APPENDIX 2

Changes in Smoothed Proportions as New Assessments New Assessments are Completed

Every time an assessment is completed, a new time point is added to each of the sequences of population proportions within the smoothing cells. This means that, even though robust/resistant procedures are used, the addition of a new point may somewhat change the values of smoothed proportions for prior years. Additionally, any changes in methodology will impact the estimates.

This means that the smoothed proportions, with the addition of the next assessment data, are apt to differ somewhat from the corresponding smoothed proportions without the new data. National Assessment has adopted the philosophy that the smoothed proportions, based on all currently available data using the best available algorithm, are the best available data. Therefore, all subsequent analyses, for any year, will be done using this best available information, even though this may produce estimates which slightly differ from prior values (before implementation of the newest smoothed proportions).

In late 1982, newly smoothed proportions were estimated for the period 1972-73 to 1981-82. These proportions are shown in Exhibits 2.4A, B and C for three ages. They were used for all analyses involving the 1981-82 assessment of citizenship/social studies, mathematics and science. It should be noted that some of the trend lines for cells in Exhibits 2.3A, B and C are quite different from those for cells in Exhibits 2.4A, B and C. Additional work may be required to obtain satisfactorily smoothed proportions for the entire time period, 1969-70 to 1981-82.

APPENDIX 2

EXHIBIT 2.4A

Smoothed Proportions by Smoothing Cell and Year
for Nine-Year-Olds from 1972-73 to 1981-82

CELL	1	2	3	4	5	6
RACE	WHITE	WHITE	WHITE	BLACK	BLACK	OTHER
REGION	ALL	ALL	ALL	SE	NOT SE	ALL
CS	BC+FR	MC	SP	ALL	ALL	ALL
YEAR						
72-73	.29556	.13224	.37592	.05497	.07676	.06465
73-74	.29807	.12530	.37837	.05570	.07726	.06540
74-75	.30059	.11836	.38081	.05642	.07777	.06615
75-76	.30311	.11142	.38325	.05714	.07828	.06690
76-77	.30563	.10447	.38570	.05787	.07878	.06765
77-78	.30815	.09753	.38814	.05859	.07929	.06840
78-79	.31066	.09059	.39059	.05932	.07979	.06915
79-80	.31318	.08365	.39303	.06004	.08030	.06990
81-82	.31822	.06977	.39792	.06149	.08131	.07140

APPENDIX 2

EXHIBIT 2.4B

Smoothed Proportions by Smoothing Cell and Year
for Thirteen-Year-Olds from 1972-73 to 1981-82

CELL	1	2	3	4	5	6
RACE	WHITE	WHITE	WHITE	BLACK	BLACK	OTHER
REGION	ALL	ALL	ALL	SE	NOT SE	ALL
CS	BC+FR	MC	SP	ALL	ALL	ALL
YEAR						
72-73	.29977	.11234	.40190	.05530	.06810	.06260
73-74	.30298	.11076	.39786	.05480	.07020	.06340
74-75	.30620	.10918	.39382	.05431	.07229	.06420
75-76	.30942	.10760	.38978	.05382	.07438	.06500
76-77	.31263	.10602	.38574	.05332	.07648	.06580
77-78	.31585	.10445	.38170	.05283	.07857	.06660
78-79	.31907	.10287	.37766	.05234	.08066	.06740
79-80	.32228	.10129	.37362	.05184	.08276	.06820
81-82	.32872	.09814	.36555	.05086	.08694	.06980

APPENDIX 2

EXHIBIT 2.4C

Smoothed Proportions by Smoothing Cell and Year
for Seventeen-Year-Olds from 1972-73 to 1981-82

CELL	1	2	3	4	5	6
RACE	WHITE	WHITE	WHITE	BLACK	BLACK	OTHER
REGION	ALL	ALL	ALL	SE	NOT SE	ALL
CS	BC+FR	MC	SP	ALL	ALL	ALL
YEAR						
72-73	.32944	.11852	.39714	.047036	.059344	.04842
73-74	.33169	.11815	.39235	.047148	.061512	.04904
74-75	.33395	.11778	.38757	.047261	.063679	.04966
75-76	.33620	.11742	.38278	.047373	.065847	.05028
76-77	.33846	.11705	.37799	.047485	.068015	.05090
77-78	.34072	.11668	.37320	.047598	.070182	.05152
78-79	.34297	.11631	.36842	.047710	.072350	.05214
79-80	.34523	.11594	.36363	.047823	.074517	.05276
81-82	.34479	.11358	.34904	.047367	.077735	.06740

APPENDIX 3

GLOSSARY OF NATIONAL ASSESSMENT TERMINOLOGY

ACCEPTABLE RESPONSE - Any response to an exercise that demonstrates achievement of the objective measured by that exercise.

ADMINISTRATION MODE - The way in which an exercise is administered. Exercises are administered in either group or individual mode. See "INDIVIDUAL ADMINISTRATION" and "GROUP ADMINISTRATION."

ADMINISTRATION TIMETABLE - The time periods during a school year when the various age groups are assessed. The time periods are:

October-December	13-year-olds
January-March	9-year-olds
March-May	17-year-olds (in-school)
Spring-Summer	17-year-olds (out-of-school)

The assessment time period for young adults, ages 26-35 has varied from year to year.

ACHIEVEMENT CLASS VARIABLE (ACV) - A classification of a respondent into performance quartiles based on his or her performance on achievement items in a particular booklet. The definition of an achievement item is developed for each assessment year and may differ slightly from year to year. However, in general an achievement item is any cognitive item for which there is a response that is keyed as correct.

AFFECTIVE EXERCISES - Attitude, experience and other items which lack an "acceptable" response or are considered to be important aspects of a learning area, but not appropriate for inclusion in knowledge or skill summaries.

AGE CLASS, AGE GROUP OR AGE LEVEL - One of the age levels sampled in the assessment. Each group is assigned a numeric value and an alphabetic designation, as follows:

Age class 1 = N = 9-year-olds
Age class 2 = T = 13-year-olds
Age class 3 = S = 17-year-olds
Age class 4 = A = Young Adults (ages 26-35)

APPENDIX 3

AGE-ELIGIBLE - Any person meeting the age definition for an assessment. Birthdate ranges for each age group in the 1969-70 to 1979-80 assessments were:

Assessment	Age 9	Age 13	Age 17	Adult
1969-70	1960	1956	10/51-9/52*	7/33-6/43
1970-71	1961	1957	10/53-9/54*	4/35-3/45
1971-72	1962	1958	10/54-9/55*	4/36-3/46
1972-73	1963	1959	10/55-9/56*	1937-1946
1973-74	1964	1960	10/56-9/57	1938-1947
1974-75	1965	1961	10/57-9/58	**
1975-76	1966	1962	10/58-9/59	**
1976-77	1967	1963	10/59-9/60	1941-1950
1977-78	1968	1964	10/60-9/61	**
1978-79	1969	1965	10/61-9/62	**
1979-80	1970	1966	10/62-9/63	**

Notes:

* In the first 4 assessments, out-of-school 17-year-olds could be up to 12 months older than shown in the table if they were not enrolled in school during March of the year prior to the assessment.

** Adult assessments were not conducted in these years.

ASSESSMENT - The documentation of the progress in knowledge, skills and attitudes of American youth. Measures are taken at periodic intervals for each learning area, with the goal of determining trends and reporting the findings to the public and to the education community.

ASSESSMENT YEAR - Annual assessments are numbered sequentially, starting with the 1969-70 assessment. Since the 1980-81 school year, it has been necessary to change to a biennial data collection due to budget constraints. BACKGROUND QUESTIONS - Several types of background questions are included in exercise booklets. A common set of questions about educational materials in the home, level of parental education, etc., have been included on the last page of the exercise booklet in every assessment. Beginning in 1975-76, 17-year-olds have been asked additional background and demographic questions, some of which are derived from the National Longitudinal Study of the High School Class of 1972. Finally, learning area-specific background questions are sometimes included.

APPENDIX 3

BACKGROUND QUESTIONNAIRE - This term is used for the form used to collect background information from out-of-school 17-year-olds and young adults.

BACKGROUND VARIABLE - See "VARIABLE."

BOOKLET - Items (exercises) are presented to respondents in booklets. Booklets are designed to be scored by optical scanning machines. Each booklet contains instructions on answering items, practice items, the assessment items and background questions. Each booklet contains approximately 30-35 minutes of assessment items and 10-15 minutes of introductory material and background questions. A booklet typically includes exercises of varying difficulty from different objectives of the learning area(s) being assessed. If more than one learning area is being assessed, the booklet may contain exercises from more than one learning area. The terms "booklet" and "package" are synonomous and may be modified by any of the following without change in meaning: assessment, exercise, item or respondent.

BOOKLET WEIGHT - See "WEIGHT."

CATEGORY (SCORING) - A classification of a response to an open-ended exercise. See "SCORING GUIDE."

CATEGORY WITHIN A VARIABLE - A subclassification within a variable. For example, male and female are categories of the variable sex.

CENSUS DIVISION - For most assessments, respondents are classified according to U.S. census divisions, based on location of their school or home (for household surveys of adults and out-of-school 17-year-olds). The categories are:

1 = New England	5 = South Atlantic
Connecticut	Delaware
Maine	District of Columbia
Massachusetts	Florida
New Hampshire	Georgia
Rhode Island	Maryland
Vermont	North Carolina
	South Carolina
	Virginia

APPENDIX 3

West Virginia

2 = Middle Atlantic
- New Jersey
- New York
- Pennsylvania

3 = East North Central
- Illinois
- Indiana
- Michigan
- Ohio
- Wisconsin

4 = West North Central
- Iowa
- Kansas
- Minnesota
- Missouri
- Nebraska
- North Dakota
- South Dakota

6 = East South Central
- Alabama
- Kentucky
- Mississippi
- Tennessee

7 = West South Central
- Arkansas
- Louisiana
- Oklahoma
- Texas

8 = Mountain
- Arizona
- Colorado
- Idaho
- Montana
- Nevada
- New Mexico
- Utah
- Wyoming

9 = Pacific
Alaska
California
Hawaii
Oregon
Washington

COGNITIVE EXERCISE - An item measuring behaviors in the cognitive domain of Bloom's Taxonomy.

COMMUNITY SIZE - Also called derived size of community (DOC). Schools (and households in the adult assessments) are classified in four categories using Census, ZIP Code, and atlas information:

1 = Big City: 1970 population of 200,000 or more
2 = Big City Fringe: Schools/segments around big cities (in the 1970 Census Urbanized Area of a Big City)
3 = Medium City: 1970 population between 25,000 and 200,000 and not in 2 above
4 = Small Places: All other places

Out-of-school 17-year-olds sampled from school lists of dropouts and early graduates receive the same community size classification as the school received in the in-school assessment. Note that this variable is not identical to either the size of community sample stratification variable or the principal's questionnaire size of community served variable.

CONTENT AREA - See "LEARNING AREA."

DELTA P-VALUE - NAEP terminology for the difference in percentages between a reporting group (such as males) and the whole nation on a particular response option or category (usually the acceptable response).

DERIVED VARIABLE, DERIVED EXERCISE PART, DERIVED VALUE - Data derived by combining responses to two or more variables or by combining values of one variable.

APPENDIX 3

DIFFICULTY LEVEL - The percentage of acceptable responses to an exercise.

DS OR DISTRICT SUPERVISOR - An individual employed to manage the in-school data collection. Each DS works in a specific section of the country contacting schools and making arrangements for the assessment. The DS also administers booklets to respondents.

DOC OR DERIVED SIZE OF COMMUNITY - See "COMMUNITY SIZE."

EA OR EXERCISE ADMINISTRATOR - An individual hired locally and trained by the district supervisor to administer booklets to in-school respondents.

ELIGIBLES - Individuals who meet the criteria for inclusion in the assessment. See "INELIGIBLES."

EXERCISE - A task designed to measure an objective. In some cases the task is expressed by a single question, in other cases the task is expressed by several related questions. Some tasks are direct measures of performance, e.g., playing a musical instrument or writing a letter. Others are indirect measures, e.g., answering a multiple choice question or writing a short answer to a question. Because NAEP does not administer "tests" but surveys educational achievement over time, the term exercise is often used instead of the terms item or test item. Item and exercise are used synonomously in the NAEP documentation.

EXERCISE BOOKLET - See "BOOKLET."

EXERCISE PART - See "ITEM PART."

EXERCISE VARIABLE - See "VARIABLE."

EXERCISE POOL - The entire set of exercises prepared for a learning area. This set includes the exercises from previous assessments used for measuring change, exercises developed for previous assessments but not used due to packaging or budgetary constraints, and newly developed exercises.

APPENDIX 3

FIELD TRIALS (TRYOUTS) - A pretest of exercises to obtain information regarding clarity, difficulty levels, timing, feasibility and special administration problems needed for revision and selection of exercises to be used in the assessment..

FOIL - See "RESPONSE OPTIONS."

FOLLOW-UP - Special studies conducted in 1972-73, 1975-76, 1976-77 and 1978-79 to follow-up a subsample of in-school 17-year-old nonrespondents to determine enrollment status and administer assessment packages, if possible. The 1972-73 study was a major study of 17-year-old nonresponse, sometimes referred to as the "No-Show Study."

FORCED NO RESPONSE - A machine generated no response to a question which is inappropriate for the respondent to answer. For example, if a respondent answers no to a question, "Did you go to high school?" then the respondent receives a forced no response to a subsequent question "What was the primary emphasis of your high school courses?".

GROUP ADMINISTRATION - The mode of administration where an exercise booklet is administered to a group of respondents in a school and a paced audio tape is used to provide uniform instructions and oral presentation of exercises. From 1969-70 to 1974-75, the target group administration size was 12. Since then session sizes have been allowed to vary, with an average target size of 16.

HAND SCORING - The rating or categorizing of responses for optical scanning. Multiple-choice exercises can be directly machine scored; however, open-ended exercises and individual exercises must be coded in scoring ovals so that they can then be machine scored. See "SCORING GUIDE."

I DON'T KNOW - The last response option on NAEP multiple-choice exercises. On open-ended exercises, respondents are encouraged to write "I don't know" if they do not know the answer.

ID NUMBER - Any identification number. Usually refers to the unique number assigned to each respondent. NAEP does not keep records of the names of any individuals.

APPENDIX 3

IN-SCHOOL (IS) ASSESSMENT - The administration of assessment booklets to 9-year-olds, 13-year-olds and 17-year-olds currently enrolled in public or private schools.

INDIVIDUAL ADMINISTRATION - The mode of administration where a booklet is administered to one person at a time. The administrator reads the exercises, using an interview format. Although there are no set response times for the individual exercises in the booklet, the booklets are designed to require no more than a total of 45 to 50 minutes of a respondent's time.

INDIVIDUAL EXERCISE - An exercise that is administered to only one person at a time. These are generally exercises that require interview techniques or active responses (such as singing).

INDIVIDUAL BOOKLET - A booklet composed of individual exercises that is administered to only one person at a time.

INELIGIBLES - Individuals who do not meet the criteria for inclusion in the assessment. Ineligibles include those who fall outside the age range, those who are not enrolled in public or private schools at the time of assessment for the 9-, 13- and in-school 17-year-old assessments. and those who are functionally handicapped so that they cannot participate in the assessment, currently defined as:

1. Non-English speaking persons;

2. Respondents identified as nonreaders during the assessments;

3. Persons physically or mentally handicapped, including educable mental retarded (EMR), in such a way that they could not respond to NAEP exercises as they are normally administered;

4. Students attending public and private schools established for the physically and/or mentally handicapped.

ITEM - See "EXERCISE."

APPENDIX 3

INVALID EXERCISE - An exercise which is technically deficit and omitted from analyses. Typical problems include ambiguous wording, multiple correct response choices and cuing by other exercises in the same booklet. Open-ended exercises which elicit many vague or irrelevant responses are included in this category.

ITEM BOOKLET - See "BOOKLET."

ITEM PART - Some items have more than one part. Generally, each part of an item asks a separate question. Parts may all pertain to one stimulus, such as a graph or a table, or may concern the same topic. Exceptions are open-ended items that ask for multiple responses ("Give three reasons...."), responses that are scored on multiple dimensions and complex individually-administered exercises.

ITEM VARIABLE - See "VARIABLE."

JACKKNIFE - The name of the algorithm used by NAEP to estimate standard errrors of percentages and other statistics. See Appendix 5 for details and references.

KEY - An indicator of which value in the response range is considered a correct or acceptable response to the exercise.

LEARNING AREA - One of the 10 areas assessed by the NAEP project: Art, Career and Occupational Development, Citizenship, Literature, Mathematics, Music, Reading, Science, Social Studies and Writing. Also called "subject area."

MARKER EXERCISE - An exercise which occurs in two or more booklets for an age group. The most common example is curriculum-related questions that appear in all booklets.

MIGRATION REGION - From 1972-73 to 1979-80, 13- and 17-year-old respondents were asked where they lived on their birthday four years earlier. (Adults were asked similar questions for various time points.) To meet confidentiality requirements, the data were recoded to show whether the respondent lived in the same state, same region, a different region (or other) compared to four years earlier.

APPENDIX 3

MODAL GRADE - The grade in which the majority of each in-school age group is enrolled. The modal grades are:

Age	Modal Grade
9	4
13	8
17	11

MULTIPLE-CHOICE EXERCISE - An exercise (item) with preprinted response options, as opposed to open-ended (or free-response or interview) exercises, where a written or verbal response is required.

N-COUNT - A numeric count. Usually refers to counts of respondents.

NAEP NUMBER - An exercise identification number which usually contains coded information about the objective and/or the content area the exercise measures.

OBJECTIVE - A desirable educational goal agreed on by scholars in the learning area, educators, and concerned lay persons. Objectives are established through a consensus approach.

OBJECTIVES REDEVELOPMENT - After the initial assessment of a learning area, one of the first steps in preparing for subsequent assessments of the learning area is a review of the learning area objectives from the previous assessment by scholars in the field, educators, and concerned lay persons. These reviews may result in revision, modification or total rewriting of the learning area objectives to reflect current curricular goals and emphases; they may also result in the endorsement of the existing objectives.

OPEN-ENDED EXERCISE - A non-multiple-choice exercise that requires some type of written or oral response.

OUT-OF-SCHOOL (OOS OR OS) ASSESSMENT - The administration of assessment materials to 17-year-olds not enrolled in secondary school. From 1969-70 to 1972-73, out-of-school 17-year-olds were located and assessed during the household survey of young adults. The yield of 17-year-olds was so low that, in 1970-71, a supplementary frame assessment was implemented, using lists of dropouts and early graduates provided by schools participating in the in-

school assessment. Since 1973-74, the school lists of dropouts and early graduates have been the only source of out-of-school 17-year-olds.

OVERLAP EXERCISE - An exercise administered to more than one age class in the same assessment.

P-VALUE - NAEP terminology for the percentage of responses to a response option or category (often the acceptable response option).

PACED AUDIO TAPE - A tape recording that accompanies group administration booklets to assure uniformity in administration. Instructions and exercises are read by the announcer on the tape so that reading difficulties will not interfere with an individual's ability to respond. An exception is assessments of reading in which the instructions are read by the announcer while the reading passages, questions and response choices are read by the respondent. Response time in included on tape. See "GROUP ADMINISTRATION."

PACKAGE - See "BOOKLET."

PACKAGE WEIGHT - See "booklet weight" under "WEIGHT."

PACKAGING PLAN - The process of allocating exercises selected for the assessment into various booklets and arranging the exercises within booklets. The plan considers exercises which should not be packaged together, exercise order, objective/content area coverage, exercise formats, difficulty, etc.

PRIMARY SAMPLING UNIT (PSU) - First stage sampling units, typically a county or group of contiguous counties.

PRINCIPAL'S QUESTIONNAIRE - A data collection form given to school principals. It contains questions about enrollments, size of community, occupational composition of community, etc. See also "SUPPLEMENTARY PRINCIPAL'S QUESTIONNAIRE."

PROBE - A small scale assessment of a specific topic or area, usually administered to only one age group. Probes are generally not designed to measure change.

PSU - See "PRIMARY SAMPLING UNIT."

APPENDIX 3

RACE - NAEP collects racial/ethnic data by visual observation of respondents (and since 1975-76, 17-year-olds have also been asked to classify themselves). Black-white-other respondent data are available for all assessments, while Hispanic respondent data are available from 1971-72 on. Since the 1976-77 adult and 1977-78 in-school assessments, standard federal categories have been used (Hispanic heritage, black, white, Asian, American Indian and Alaskan Native). One inconsistency occurs in race classification for the 1970-71 assessment. In that year Hispanic respondents are classified as whites, in other years Hispanic respondents are classified as other.

RECEIPT CONTROL - Procedures implemented by scoring staff to check in and screen materials from the field. Information gained from receipt control procedures is relayed to field administration staff so that any errors may be corrected.

RECYCLED EXERCISES - See "SECURE EXERCISES."

REGION - NAEP uses the geographic region definition of the office of Business Economics, Department of Commerce, for both sample stratification (1970-71 on) and reporting. Respondents are classified by region based on the location of their school or home (for household surveys of out-of-school 17-year-olds and adults). The categories are:

1 = Northeast:

Connecticut	New Hampshire
Delaware	New Jersey
District of Columbia	New York
Maine	Pennsylvania
Maryland	Rhode Island
Massachusetts	Vermont

2 = Southeast:

Alabama	Mississippi
Arkansas	North Carolina
Florida	South Carolina
Georgia	Tennessee
Kentucky	Virginia
Louisiana	West Virginia

3 = Central:

Illinois	Missouri
Indiana	Nebraska
Iowa	North Dakota

APPENDIX 3

Kansas	Ohio
Michigan	South Dakota
Minnesota	Wisconsin

4 = West:

Alaska	Nevada
Arizona	New Mexico
California	Oklahoma
Colorado	Oregon
Hawaii	Texas
Idaho	Utah
Montana	Washington
	Wyoming

RELEASE NUMBER - An identification number assigned to an exercise when NAEP reports are organized. The reporting theme to which the exercise belongs is coded in the number.

RELEASED EXERCISE - An exercise for which results and exercise text have been reported to the public. This type of exercise is in the public domain and may be used by anyone. However, if the exercise includes copyrighted material a user must obtain permission from the copyright holder to use that material.

RELEASED EXERCISE SET - A set of the released exercises, including documentation and scoring guides, made available to state and local education agencies, the research community and the general public.

REPLICATE - Replicate is used in two distinct ways by NAEP. For sample selection and variance estimation purposes, the variance estimation replicate is a first-stage sampling unit in the 1975-76 and subsequent assessments. Except for very large SMSAs (which are further stratified into multiple variance estimation replicates), variance estimation replicates correspond to primary sampling units. For purposes of allocating workload (called package assignment), a package assignment replicate is a set of schools in which one group administration of every exercise booklet is scheduled. Variance estimation replicates contain one to three package assignment replicates.

REPORTING GROUPS - Categories of variables for which National Assessment data are reported. They typically include

sex, race, region, community size, type of community, modal grade and parental education.

RESCORE - If an open-ended exercise is affected by recent history, it may be necessary to make changes in the scoring categories of the exercise when it is used to measure change and then to rescore the responses from the previous assessment using the new scoring categories. For other open-ended exercises, it is necessary to have responses from both assessments scored simultaneously to insure comparability of scoring.

RESPONDENT - A person who responds to the exercises and background questions in an assessment booklet. In-school assessment respondents respond to only one booklet; out-of-school 17-year-olds and adults may respond to as many as three booklets.

RESPONDENT WEIGHT - See "WEIGHT."

RESPONSE OPTIONS - Different alternatives to a multiple-choice question that can be selected by the respondent. Also called "foils."

RESPONSE TAPE - Audio tape recording of a verbal or musical response made by the respondent.

REVIEW CONFERENCE - A conference held to review the objectives of a learning area to assure their acceptance by scholars, educators, and lay persons or to review exercises for racial, ethnic, social or regional bias and to assure their acceptance as measures of the objectives by scholars, educators and lay persons.

SAMPLE - A subset of the population who are assessed in order to estimate the performance of the total population.

SCHOOL WEIGHT - See "WEIGHT".

SCORING GUIDE - A guide for hand-scoring an open-ended exercise that specifies descriptive or diagnostic categories by giving definitions and example responses. Categories are usually defined as "acceptable" or "unacceptable."

SCORING OVALS - Scannable ovals printed beside multiple-choice foils and printed at the bottom of the page for

open-ended exercises (to be used in hand scoring). When the ovals are marked, they can be machine scored.

SECURE EXERCISES - The set of exercises that is kept secure from one assessment to the next and used to measure changes (growth or decline) in performance for the learning area. Also called unreleased exercises and recycled exercises.

SIZE AND TYPE OF COMMUNITY - The seven size and type of community reporting categories consist of three "extreme" types of community and four "residual" community sizes. Each extreme category includes approximately 10% of the respondents at each age level; the remaining respondents are classified according to one of the Community Size categories. The extreme categories are:

1 = Extreme Rural - Rural areas where a high proportion of adults are farmers or farm workers and a low proportion are professional, managerial or factory workers. At least some of the respondents are from open country or places less than 2,500 population, excluding places greater than 10,000 (from the Principal's Questionnaire) and the suburbs of medium and large cities. These respondents must be located in Community Size category 4 (Small Places).

2 = Low Metro - City areas where a high proportion of the adult population is either not regularly employed or on welfare and a low proportion is employed in professional or managerial positions. These respondents must be in Community Size categories 1 or 2 (Big City or Big City Fringe).

3 = High Metro - City areas where a high proportion of adults are employed in professional or managerial positions and a low proportion are factory or farm workers, not regularly employed or on welfare. These respondents must be in Community Size categories 1 or 2 (Big City or Big City Fringe).

APPENDIX 3

Respondents were placed into one of the extreme categories based on an occupational profile of the community in which the school was located. For in-school respondents at each age and for the supplementary frame sample, the school principal provided estimates of the percentage of students whose parents fit into each of six different occupational categories. From these proportions of various occupational categories, the three extreme groups are obtained. The remaining respondents are classified into four "residual" categories according to the Community Size category in which the school or household was located:

4 = Main Big City - Areas located within the limits of big cities with population greater than 200,000 but not included in either the Low Metro or High Metro categories (Community Size category 1, less High Metro and Low Metro).

5 = Urban Fringe - 1970 Census urbanized areas of big cities of population greater than 200,000 but outside the city limits and not classified as Low Metro or High Metro (Community Size category 2, less High and Low Metro).

6 = Medium Cities - Cities with population between 25,000 and 200,000, excluding metropolitan areas of big cities (Community Size category 3).

7 = Small Places - Open country or places with populations less than 25,000, excluding those classified as Extreme Rural (Community Size category 4, less Extreme Rural).

Note: Initial analyses and reports for the 1969-70 to 1972-73 assessments employed a somewhat different definition of Size and Type of Community. Since late 1975, all analyses and reports involving those assessments have utilized the current definition.

SMSA - Standard metropolitan statistical area, an economic and social unit defined by the U. S. Census Bureau.

SOURCE DOCUMENTS - Materials containing original respondent responses such as booklets or workbooks or containing original background information from other sources such as schools.

APPENDIX 3

STANDARD ERROR - A measure of sampling variability for a statistic. Because of NAEP'S complex sample design, standard errors are estimated by jackknifing first stage sample estimates.

STANDBY SCHOOLS - Schools with too few age-eligibles to accomodate at least one group administration of a booklet are called standby schools. They are typically stratified and sampled differently than schools with more age-eligibles.

STEM - The portion of an exercise or exercise part that states the problem or asks the question.

STIMULUS - The task which is presented to the respondent. This includes one or more written or oral questions (see STEM). It may also include additional material such as a reading passage, map, charts or graph, photograph or picture, musical selection, oral reading, a physical object and so forth.

STOC - See "SIZE AND TYPE OF COMMUNITY."

SUBJECT AREA - See "LEARNING AREA."

SUBPOPULATION OR SUBGROUP - Groups, such as males and females, within the national population for which results are reported.

SUPPLEMENTARY FRAME - Used to locate out-of-school 17-year-olds, the supplementary frame consists of lists of dropouts and early graduates obtained from a subsample of schools included in that year's assessment sample. (See also "OUT-OF-SCHOOL ASSESSMENT.")

SUPPLEMENTARY PRINCIPAL'S QUESTIONNAIRE - A data collection form given to principals. They are asked to respond to questions concerning course offerings, materials, and staffing specific to the learning area(s) being assessed.

APPENDIX 3

TARGET POPULATION - Individuals who meet the criteria for inclusion in the assessment, including those falling within the age range, enrolled in public or private schools at time of the 9-, 13- and in-school 17-year-old assessments, and functionally able to participate in the assessment.

TAILSHEET - The last page of questions in an in-school assessment booklet. Respondents provide personal background information about their home environment, parents' education levels, etc.

TAPESCRIPT - A script prepared for the announcer to use in producing the paced audio tape. It indicates exactly what is to be read and the amount of response time to be allowed for each exercise. See "PACED AUDIO TAPE."

TIMING - Most NAEP exercises are administered with a paced audio tape to standardize data collection conditions, including the amount of time allowed to respond to each exercise.

TOC OR TYPE OF COMMUNITY - The three extreme Size and Type of Community categories.

UNIVERSAL - A classification category which identifies exercise parts included in an overall summary of a content area, for example a summary of citizenship/social studies performance. In some cases it identifies exercise parts that are included in any type of summary for an area, for example summaries of cognitive science performance and affective science performance.

UNRELEASED EXERCISE - See "SECURE EXERCISE."

APPENDIX 3

VARIABLE - Any data field on National Assessment data files. Different conventions and documentation standards apply to exercise (item) variables than to background (including demographic and housekeeping) variables. NAEP documents often use the term variable as if it applied only to reporting variables, such as region, sex, and race.

WEIGHT OR SAMPLING WEIGHT - National Assessment uses deeply stratified, multistage probability sample designs with differential sampling rates for various subpopulations. The selection probability of each respondent is calculated and its reciprocal (adjusted for nonresponse) is used to weight each response in statistical calculations. The weights compensate for unequal sampling rates and insure proper representation in the population structure. National Assessment data files include several distinct types of weights, which differ primarily in the way nonresponse adjustments are made. The booklet (or respondent or student) weight is the reciprocal of the probability of the respondent being selected for a particular exercise booklet, adjusted for nonresponse to that booklet. This weight is used by National Assessment in virtually all statistical analyses. The multi-booklet analysis weight is the reciprocal of the probability of selecting a respondent for any booklet, adjusted for nonresponse to all booklets. The multi-booklet analysis weight can be used for analyses across all booklets, but the booklet weight works equally well for either booklet-specific or cross-booklet analyses. Beginning with the 1979-80 assessment, an additional school-based weight has also been included. This weight is the inverse of the probability of selecting a school, adjusted for school nonresponse. It is used for school-level analyses where the school is considered the unit of analysis.

WORKBOOK - In some assessments (1972-73 and all art assessments), respondents were asked to answer certain items in a separate booklet, called a workbook.

APPENDIX 4

INFERENCE FROM SURVEY DATA AND ANALYSIS OF DATA FROM COMPLEX SAMPLE SURVEYS

NAEP is a sample survey which collects data by using a multistage design with unequal probabilities of selection of elements. Such a survey design is commonly referred to as "complex" sampling. It permits the collection of representative data for the population and many subgroups in an extremely cost effective manner. NAEP data are suited for various descriptive purposes.

1. Estimation of achievement by exercise and summary level for the age populations as a whole and for a variety of subgroups of the populations.

2. Estimation of changes in achievement on a variety of tasks (and summaries) over time. Change can be measured for populations and for subgroups.

3. Explorations of observed associations between variables of interest.

4. The various descriptive statistics should be computed taking the structure of the sample into account.

The NAEP data present certain difficulties in analysis, which are shared by all complex surveys. These difficulties include problems in inference due to the type of data and complications in analysis due to the sample design. These difficulties are detailed in the following sections.

APPENDIX 4

Inference From Survey Data

Because the data collected from a survey are observational and do not result from the control of variables (as do data collected from a designed experiment) they require careful interpretation. In experimental research it is often possible to infer causation because the researcher has controlled the individual variables. In surveys, because the values of the various variables are not controlled and because the variables tend to be related in value to each other and to other extraneous variables, such control is generally not possible. Because correlation, in itself, does not prove causation, the direction of causation is often unclear for many variables from survey data, such as behaviors and correlated attitudes.

In applying standard statistical tests, it is often the case that undesirable confounding occurs between pertinent variables and other, perhaps extraneous, variables; hence, tests performed in the presence of these, often unremovable, confounding variables require very careful interpretation. Because of this, NAEP data should be examined in an exploratory, rather than confirmatory, framework.

Complications in Analysis Due to Sample Design

Complications in analysis occur because of the multistage design of the NAEP sample. Many standard statistical procedures assume that data are acquired by means of a simple random sample of the population and that individuals are independent. Because the NAEP sample employs stratification, clustering and unequal probabilities of selection, these assumptions are not met.

Certain subgroups of the population are sampled at a higher rate than the remainder of the population in order to ensure adequate representation. Consequently, those subgroups, which tend to have different characteristics (including achievement) than the remainder of the population, are over-represented in the sample. Analyses which ignore this are apt to produce biased and misleading information, since those groups may have unwarranted impact. This difficulty is avoided by conducting weighted analyses, in which the weight assigned to an individual is related to the reciprocal of his or her probability of selection.

APPENDIX 4

Because of cost and administrative efficiency considerations, NAEP data are obtained by selecting a number of schools and then selecting a number of students within each of the schools. Since the students are selected in clusters, observations from various students are not independent. Student responses within a school tend to be relatively more homogeneous than student responses from different schools.

Ignoring the effect of stratificiation and clustering in analysis tends to produce severe underestimates of the variablity of statistics. Many studies (such as Ross (1976), Kish and Frankel (1974), Frankel (1971)) have demonstrated large influences of complex survey designs on sampling errors of various statistics, such as regression and correlation coefficients. It has been demonstrated (Shah, Holt and Folsom (1977)) that regression analyses of data from complex survey samples produce tests of significance which are generally too liberal when the structure of the sample is not taken into account. Additionally, as noted by Fellegi (1979) in his consideration of goodness of fit tests, the distribution of certain statistics, as well as their dispersions, can be affected by the sample design.

Because of the nonlinearity of many of the statistics of interest, it is not currently possible to exactly account for the sample structure in analysis. However, several procedures exist which approximately do this. Among these are:

1. jackknifing--the procedure used by NAEP (detailed by Folsom (1977));

2. balanced repeated replications (detailed by McCarthy (1969)); and

3. Taylor series approximation (detailed by Folsom (1977)).

Shah, Holt and Folsom (1977) give procedures for estimation and hypothesis testing of regression models for data from a complex sample survey using Taylor series approximations. A general procedure for obtaining approximate variances by Taylor series approximations is given by Woodruff and Cansey (1976).

APPENDIX 4

Fellegi (1979) and McCarthy (1969) give procedures for using balanced repeated replications to conduct goodness of fit tests.

Folsom (1977) gives the procedures used by NAEP in its analyses using jackknifing methodology.

A comparison of the performance of the three procedures for means, variances, correlations, and regression coefficients is given by Frankel (1971) and Kish and Frankel (1974).

Approximately Accounting for Sample Design with Design Effects

It may be possible to approximately account for the effects of the sample design by using an inflation factor, the design effect, developed by Kish (1967) and extended by Kish and Frankel (1974). The design effect for a statistic is the ratio of the actual variance of the statistic (taking the sample design into account) over the variance assuming a simple random sample with the same number of elements. The design effect may be used to adjust error estimates based on simple random sampling assumptions so that the effect of the design is approximately accounted for. In practice, this is often accomplished by dividing the total sample size by the design effect and then using this effective sample size in the computation of errors. It must be kept in mind, however, that the value of the design effect depends on the variables considered in a particular analysis as well as the clustering effects occurring among sampled elements.

1. Based on empirical results and theoretical considerations, Kish and Frankel (1974) have developed several conjectures about design effects:

2. Generally, the design effects for complex statistics from complex samples are greater than one and so variances based on simple random sampling assumptions tend to be underestimates.

3. The design effects for complex statistics (such as regression coefficients) tend to be smaller than the corresponding design effects for means of the same variables. Hence, the latter estimates, which are more easily computed, tend to give over-estimates of the design effects of complex statistics.

4. The design effects of complex statistics tend to resemble those of means, variables with a high design effect for the mean tend to also have high design effects for complex statistics involving those variables.

National Assessment has computed design effects for various types of statistics. A key statistic that has been examined is the estimated performance, P, of a subgroup of the population on an assessment exercise. This estimate, which is a weighted mean of the responses of individuals in the subgroup to the exercise, has a design effect of the form

$$deff(P) = Var(P)/(P(1-P)/N).$$

In the above, N is the total number of sampled individuals in the subgroup who responded to the exercise and Var(P) is the jackknifed variance of P (which takes the sample design into account). Upon examination of the distributions of design effects across exercises within reporting category (i.e., whites, low metropolitan, northeast) for a variety of ages, subjects and years, two things were noted:

1. The various distributions were remarkably similar, although there were relatively more large design effects for the smaller, more clustered, subgroups such as low metropolitan.

2. The centers of all distributions were close to two, with the majority of values being less than two.

Because of this, a design effect of two was deemed a reasonable value to use for complex analyses. Based on Kish and Frankel's conjectures, this value should be an approximate upper bound for design effects of complex statistics. In a small scale regression analysis on NAEP data, the mean value of the design effects of regression coefficients was, in fact, two.

APPENDIX 4

National Assessment has conducted regression analyses using design effects to account for the sample design and has found them adequate for exploratory purposes. The procedure used by National Assessment for regression analysis (called standard regression) using standard statistical packages (such as SPSS, BMDP and SAS) is as follows:

1. Every individual has a sampling weight which gives the relative importance of that individual in the population. These weights are necessary to provide proper contributions of the various subgroups in the estimates.

2. Each individual is assigned a scaled weight which is the person's sampling weight times the factor f = (N/2)WTOT. In the formation of f, N is the sample size, N/2 is the effective sample size, and WTOT is the total of the sampling weights of all N individuals in the analysis. The scaled weights sum to the effective sample size and maintain the relative weightings of the individuals.

3. A weighted regression analysis is then conducted using scaled weights.

The estimates of the regression parameters using this procedure are identical to those obtained using the more complicated Taylor series and jackknifing procedures given by Folsom (1977). Although the estimated variances of the parameter estimates obtained using the above procedure differ from the more exact variance estimates (from Taylor series and jackknifing), they tend to be similar and appear adequate for exploratory purposes. However, significance tests using the standard regression approach need to be modified. The actual number of degrees of freedom for error is approximately the number of variance estimation replicates minus the number of strata rather than the effective sample size minus the number of parameters, the value given by the standard analysis. Therefore, significance should be assessed by comparison to critical values with the smaller error degrees of freedom.

APPENDIX 4

It should also be noted that the estimated covariances, between regression parameters using the standard regression can differ substantially from those estimated using Taylor series and jackknifing. This is because the joint distributions of subgroup performance and subgroup size are accounted for in the Taylor series and jackknifing procedures while they are not in the standard regression.

Tests of goodness of fit and independence can be performed on NAEP data by using the design effect approach. To accomplish this, the counts of individuals within various cells are replaced by sums of scaled weights where, again, the sum across all subjects of the scaled weights is the effective sample size. Apart from the substitution of scaled weight totals, the analyses proceed in the usual manner. Felegi (1979) has empirically investigated this approach and found that although the tests are somewhat liberal this approach peforms acceptably well for survey designs such as National Assessment's.

References

Fellegi,I.P. "Approximate Tests of Independence and Goodness of Fit Based on Stratified Multi-Stage Samples." Survey Methodology, VOL. 4, 1979.

Folsom, Ralph E. "National Assessment Approach to Sampling Error Estimation," Sampling Error Monograph 250-796-5. Research Triangle Institute, Research Triangle Park, North Carolina, 1977.

Frankel, Martin R. Inference from Survey Samples. Institute for Social Research, University of Michigan, Ann Arbor, Michigan, 1971.

Kish, Leslie. Survey Sampling. New York: John Wiley and Sons, 1967.

Kish, Leslie and Martin R. Frankel. "Inference from Complex Samples," Journal of the Royal Statistical Society Series B, vol. 36, 1974.

APPENDIX 4

McCarthy, P. J. "Pseudo-Replications: Half Samples," Review of the International Statistical Institute, vol. 37, 1969.

Ross, Kenneth Norman. Searching for Uncertainty. Occasional Paper no. 9. Hawthorn, Victoria, Australia: Australian Council for Educational Research, 1976.

Shah, Babubhai V., Mary Margaret Holt, and Ralph E. Folsom. "Inferences about Regression Models from Sample Survey Data." Paper presented at the International Association of Survey Statisticians Third Annual Meeting, New Delhi, December 5-15, 1977.

Woodruff, Ralph S. and Beverley D. Causey. "Computerized Method for Approximating the Variance of a Complicated Estimate." Journal of the American Statistical Association, vol. 71, 1976.

APPENDIX 5

NATIONAL ASSESSMENT ESTIMATION OF STANDARD ERRORS

Several measures of achievement that National Assessment uses in its reports are described in the Data Analysis section of this document. The sample design is a complex, deeply stratified, multistage probability sample design. A reasonably good approximation of standard error estimates of these achievement measures can be obtained by applying the jackknife procedure to first-stage sampling units (replicates) within strata, using the method of successive differences and accumulating across strata.

In this appendix, the measures of achievement are first defined in algebraic form, followed by a description of the jackknife method used by National Assessment to estimate their standard errors.

Measures of Achievement

Based on the sample design, a weight is assigned to every individual who responds to an exercise administered in an assessment. The weight is the reciprocal of the probability of selecting a particular individual to take a particular exercise with adjustment for non-response. Since the probabilities of selection are based on an estimated number of people in the target age population, the weight for an individual estimates the number of similar people that that individual represents in the age population. As explained in Appendix 2 the weights were adjusted to reflect information from previous assessments on population distributions.

A sum of the weights for all individuals at an age level responding to an exercise is an estimate of the total number of people in that age population. A sum of weights for all individuals at an age responding correctly to an exercise is an estimate of the number of people who would be able to respond correctly in the age population if the entire popula-

tion were assessed. These concepts also apply to any reporting group (e.g., defined by region, sex, etc.) and category of response (e.g., correct, incorrect and "I don't know").

Let W^e_{ihk} = sum of weights for respondents to exercise e who are in reporting subgroup i and who are in the kth replicate of the hth sampling stratum, and

C^{ej}_{ihk} = sum of weights for respondents to exercise e who are in reporting subgroup i, who are in the kth replicate of the hth sampling stratum, and who selected response category j (e.g., correct foil) for the exercise.

Note that $W^e_{ihk} = \sum_j C^{ej}_{ihk}$

Then, summing k over the n_h sample replicates in the stratum h, and summing over the H sampling strata,

$W^e_{i++} = \sum_{h=1}^{H} \sum_{k=1}^{n_h} W^e_{ihk}$ estimates the number of eligibles in the population who are in subgroup i.

Similarly, $C^{ej}_{i++} = \sum_{h=1}^{H} \sum_{k=1}^{n_h} C^{ej}_{ihk}$

estimates the number of eligibles in the population who are in subgroup i and who would select response category j for exercise e.

An estimate of the proportion of the eligibles in the age population in group i who would select response category j on exercise e is:

APPENDIX 5

$$(1)\quad P^{ej}_{i} = C^{ej}_{i++} / W^{e}_{i++}.$$

In the special case where the proportion of all age eligibles who would select response category j on exercise e is estimated, the index A (for ALL) will be used in place of i as follows:

$$(2)\quad P^{ej}_{A} = C^{ej}_{A++} / W^{e}_{A++}.$$

In National Assessment reports, the proportion in (1) multiplied by 100 is called the group percentage, and the proportion in (2) multiplied by 100 is called the national percentage. The difference between the proportion in subgroup i who would select category j on exercise e and the proportion in the nation is denoted by:

$$(3)\quad dP^{ej}_{i} = P^{ej}_{i} - P^{ej}_{A}.$$

National Assessment also reports the arithmetic mean of the percentage of correct responses over sets of exercises corresponding to the measures in (1), (2) and (3). These means are taken over the set of all exercises or a subset of exercises classified by a reporting topic or content objective. The mean percentage of correct responses taken over m exercises in some set of exercises corresponding to measures (1), (2) and (3) are, respectively:

$$(4)\quad \bar{P}^{}_{i} = \frac{1}{m} \operatorname{Sum}_{e} C^{e}_{i++} / W^{e}_{i++},$$

$$(5)\quad \bar{P}^{}_{A} = \frac{1}{m} \operatorname{Sum}_{e} C^{e}_{A++} / W^{e}_{i++} \quad \text{and}$$

$$(6)\quad \overline{dP}^{}_{i} = \bar{P}^{}_{i} - \bar{P}^{}_{A}.$$

Note that the response category subscript j has been suppressed since the means are understood to be taken over the correct response category for each exercise.

Each of these six achievement measures is computed and routinely used in reports describing achievement data for any assessment. The simple difference in these measures between two assessments of the same exercise (or sets of exercises) provides six measures of change in achievement that are routinely used in National Assessment's change reports. The next section describes how standard errors are estimated for the 12 statistics routinely used in NAEP reports.

Computation of Standard Errors

In order to obtain an approximate measure of the sampling variability in the statistics (1) through (6), a jackknife replication procedure for estimating the sampling variance of non-linear statistics from complex, multistage samples was tailored to National Assessment's sample design. Miller (1968, 1974) and Mosteller and Tukey (1977) provide information about the jackknife technique, while reference (1) describes how the procedure is used in estimating standard errors for National Assessment's sample designs.

To demonstrate the computational aspects of this technique, consider estimating the variance of the statistic in (1) - the proportion of age-eligibles in subgroup i who would select response category j on exercise e.

This statistic is based on the data from all the n_h replicates in the H strata. Let $p_{i\text{-}hk}^{ej}$ be defined as a

APPENDIX 5

replication estimate of p_{ij} and constructed from all the replicates excluding the data from replicate k in stratum h. These replication estimates are computed as if the excluded replicate had not responded and a reasonable nonresponse adjustment is used to replace the data in replicate hk in estimating p_{ij}. Several choices for replacing the data in replicate hk are available. In order to obtain a convenient and computationally efficient algorithm for approximating standard errors, National Assessment replaces C_{ihk}^{ej} and W_{ihk}^{e} from the hk(th) replicate with corresponding sums from another paired replicate in the same stratum. The replicate estimate is then computed. The replicate estimates to be used in the calculations are determined by arranging all the replicates in each stratum into successive pairs. That is, replicate 1 is paired with replicate 2, replicate 2 with replicate 3, 3 with 4, ... $(n_h - 1)$ with n_h and replicate n_h with replicate 1.

The contribution to the variance of p_{ij} by each pair of replicates is the change in the value of the statistic incurred by replacing the data from each replicate in the pair with the data from the other replicate in the pair and recomputing p_{ij} in the usual way. This produces two replicate estimates. Squaring the difference between these replicate estimates and then dividing by 8 measures the contribution of this pair of replicates to the total variance. The sum of these contributions over all n_h successive pairs in the stratum is the contribution by stratum h to the total variance. The square root of the sum of the H stratum contributions is the estimate of the standard error of p_{ij}.

APPENDIX 5

Algebraically, the two replicate estimates for the pair k, k+1 (where k=1,...n and n +1=1) are:
h h

$$(7)\quad P^{ej}_{i-hk} = \frac{C^{ej}_{i++} - C^{ej}_{ihk} + C^{ej}_{ih(k+1)}}{W^{ej}_{i++} - W^{ej}_{ihk} + W^{ej}_{ih(k+1)}}$$

and

$$(8)\quad P^{ej}_{i-h(k+1)} = \frac{C^{ej}_{i++} - C^{ej}_{ih(k+1)} + C^{ej}_{ihk}}{W^{ej}_{i++} - W^{ej}_{ih(k+1)} + W^{ej}_{ihk}}$$

The contribution to the total variance from stratum h is:

$$(9)\quad var\,(P^{ej}_{ih}) = \frac{1}{8}\sum_{k}^{n_h}\left(P^{ej}_{i-hk} - P^{ej}_{i-h(k+1)}\right)^2$$

And, finally, an estimate of the standard error of p^{ej}_{i} is:

$$(10)\quad SE\,(P^{ej}_{i}) = \left(\sum_{h}^{H} var\,P^{ej}_{ih}\right)^{1/2}$$

Multiplying p^{ej}_{i} by 100 yields the percentage of response to category j. Multiplying $SE(p^{ej}_{i})$ by 100 yields the corresponding estimated standard error of the percentage.

APPENDIX 5

In general, the jackknifed standard errors of the proportion estimates will be larger than the simple random sampling formula $(pq/n)^{1/2}$, where $p=p_{ej}$, $q=1-p$ and n_i is the number of sampled respondents in subgroup i who took the exercise. The larger size of $SE(p_{ej})$ reflects mainly the loss of precision due to cluster-sampling of schools and students (see Appendix 4 for further discussion of this point).

The standard errors for the achievement measures (2) through (6) are computed through a series of steps analogous to those followed in computing $SE(p_{ej})$. The most complicated step in computing standard errors occurs in forming the paired replicate estimates analogous to (7) and (8) for each successive pair of replicates. Once this bookkeeping chore is done, the computations (9) and (10) follow in a straightforward manner.

The standard errors for the differences between two assessments for any of the achievement measures (1) through (6) are computed as the square root of the sum of the squared standard errors from each of the separate assessments.

The size of the standard errors depends largely not only on the number of replicates and schools included in the sample, but also on the number of respondents in each of the reporting groups. Exhibit 5.1 shows the average proportion of students responding to an exercise booklet for each of the reporting groups for each age. The proportions in Exhibit 5.1 are for the 1979-80 assessment, averaged over booklets and smoothed over several adjacent assessments. Figures for any particular booklet can be expected to deviate somewhat from Exhibit 5.1.

The size of the standard errors of the means of the achievement measures for sets of exercises is also influenced by the number of exercises in the exercise set and the number of booklets over which the items in the set are spread.

APPENDIX 5

References

Folsom, R. E. Jr. National Assessment Approach to Sampling Estimation, Sampling Error Monograph 25U-796-5. Research Triangle Park, N.C.: Research Triangle Institute, 1977.

Miller, R. G. Jr. "A Trustworthy Jackknife," Annals of Mathematical Statistics, No. 39 (1968), pp.567-82.

Miller, R. G. Jr. "The Jackknife - A Review," Biometrika, No. 61 (1974), pp.1-15.

Mosteller, F. and J. W. Tukey. Data Analysis and Regression (Chapter 8). Reading, Mass.: Addison-Wesley, 1977.

APPENDIX 5

EXHIBIT 5.1

Reporting Groups for In-School Students Ages 9, 13 and 17

Reporting Groups	1979-1980 Age 9	Age 13	Age 17
Sex			
Male	.50	.50	.48
Female	.50	.50	.52
Race			
White	.79	.80	.83
Black	.14	.13	.12
Other	.07	.07	.05
Region			
Northeast	.25	.25	.25
Southeast	.22	.23	.20
Central	.27	.27	.29
West	.26	.25	.26
Parental education			
Not graduated high school	.09	.13	.15
Graduated high school	.24	.32	.32
Post high school	.33	.42	.48
Unknown	.34	.13	.05
Type of community			
Extreme rural	.08	.10	.08
Low metro	.07	.07	.09
High metro	.11	.11	.11
Other	.74	.72	.72
Size of community			
Big city	.20	.21	.19
Fringes around big cities	.22	.22	.26
Medium city	.12	.11	.11
Smaller places	.46	.46	.44
Grade in school			
<3, <7, 10	<.01	.02	.02
3, 7, 10	.23	.25	.13
4, 8, 11	.75	.72	.75
>4, >8, 12	<.01	<.01	.10
Other	<.01	<.01	<.01